

**UTILITY
PATENT APPLICATION**

on

**METHODS TO CONSTRUCT MULTIMERIC DNA AND POLYMERIC
PROTEIN SEQUENCES AS DIRECT FUSIONS OR WITH LINKERS**

by

Stuart J. Bussell

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

"EXPRESS MAIL" MAILING LABEL NUMBER EL 89218150705

DATE OF DEPOSIT July 16, 2003

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and is addressed to:
The Assistant Commissioner for Patents, Washington, D.C. 20231

David R. Preston
(Typed or printed name of person mailing paper or fee)

David R. Preston
(Signature of person mailing paper or fee)

Number of Figures: [23]
GNP-00101.P.1-US

David R. Preston & Associates
12625 High Bluff Drive
Suite 205
San Diego, CA 92130

METHODS TO CONSTRUCT MULTIMERIC DNA AND POLYMERIC PROTEIN
SEQUENCES AS DIRECT FUSIONS OR WITH LINKERS

CROSS REFERENCE TO RELATED APPLICATIONS

5 This application claims priority to U.S. provisional application number US
60/396,466, filed July 16, 2002, naming Stuart Bussell as inventor.

SEQUENCE LISTING

A sequence listing is provided in electronic and printed form and as an appendix
10 to this application.

BACKGROUND

The present invention relates generally to recombinant DNA technology and
recombinant protein expression, and more specifically, to constructs comprising repeat
DNA sequences and to methods of making constructs comprising repeat DNA sequences,
15 including constructs that encode polymer peptides and proteins, in which monomers are
either fused directly or with linkers.

Recombinant proteins have become an important class of therapeutics and
diagnostics since their introduction in the 1980s. The first recombinant protein
therapeutics replaced products isolated from either animal or human tissue. For example,
20 recombinant human growth hormone (recombinant human GH or rhGH) replaced
material isolated from the pituitaries of human cadavers (Jorgenson, *Endocrine reviews*
12:189, 1991). The need arose because of the transmission of a rare fatal disease, called
Creutzfeldt-Jakob disease (CJD), that is transmitted from impurities in pituitary derived
hGH. The level of control possible with the recombinant version enabled production of
25 drug certifiably free of known communicable agents.

Another example of an early recombinant protein is recombinant human insulin
(rhl) (Chien, *Drug Development and Industrial Pharmacy* 22:753, 1996). In this case,
the recombinant product replaced, or supplemented, insulin isolated from the pancreases
from swine and cattle. The recombinant protein exactly matches the one found naturally
30 in humans, in contrast with the animal versions that differ by one to three amino acids.

More recombinant protein therapeutics followed including interferons, interleukins, hematopoetic factors, monoclonal antibodies, and others.

In the diagnostic field, antibodies, both natural and engineered, are used to recognize and signal the presence of clinical markers. An advantage of engineered antibody fragments over full-length antibodies is that they are amenable to production in facile expression systems such as *E. coli* or *P. pastoris* (Pennell et al., *Res Immunol* 149:599, 1998).

Some of the *in vivo* characteristics of recombinant drugs are described by their pharmacokinetic parameters. The field of pharmacokinetics concerns itself with the absorption, distribution, metabolism, and excretion (ADME) of compounds delivered *in vivo*. Basically, pharmacokinetic parameters describe the concentration of a drug distributed throughout the body over time.

Generally, absorption of protein drugs requires delivery by injection. A body's natural barriers tend to prevent the absorption of intact proteins if any other routes of delivery are used. The digestion system breaks down proteins administered orally, while the body's various epidermal surfaces prevent absorption throughout the body.

Once injected, proteins tend to distribute throughout the circulatory system where they can react (part of metabolism) with other molecules or undergo excretion. Mathematical models, of varying complexity, are available to explain experimental measurements of drug concentrations as a function of time. One of the basic pharmacokinetic parameters is a drug's half-life, $t_{1/2}$, which is characteristic of the drug's duration in the bloodstream.

A key determinant to a protein's half-life in the blood is its size, and this is a result of elimination of proteins from the blood by glomerular filtration in the kidneys (Venkatachalam et al., *Circulation Research* 43:337, 1978). Basically, the filtration allows proteins smaller than 60 kilodaltons (kD), and other similarly sized molecules, to pass out of the blood, resulting in urinary excretion, while retaining larger ones. This has a major impact on the dosing regimen for a given protein. Proteins smaller than 60 kD tend to need daily, or more frequent, injections.

One strategy to minimize the discomfort and inconvenience of daily injections is to prolong the action of proteins once introduced *in vivo*. Two basic strategies are used.

- One involves the formulation of the protein into a slow release formulation (Putney et al., *Nature Biotechnology* 16:153, 1998). An example of this technique involves formulating proteins into a biocompatible polymer, poly lactic co-glycolytic acid (PLGA), that dissolves slowly over time, releasing protein during the dissolution process.
- 5 Recombinant hGH is one protein successfully formulated this way (Johnson et al., *Nature Medicine* 2:795, 1996). A disadvantage of this technique that complicates its widespread application is the challenge of formulating and manufacturing each protein so that it is stable during processing and use. Furthermore, injections of PLGA formulated proteins can be uncomfortable.
- 10 The other strategy to prolong a protein's *in vivo* action involves modifying the protein so that it acts like a larger particle and is excreted more slowly through the kidneys. While prolonging the proteins *in vivo* residence, the modification must avoid adverse consequences such as immunogenicity, toxicity, unwanted changes to the molecules distribution, and unwanted changes to its activity.
- 15 A common technique in protein modification involves conjugating a native protein to polyethylene glycol (PEG) or another protein (Roberts et al., *Adv Drug Deliv Rev* 54:459, 2002). PEG molecules are manufactured at all ranges of molecular weights. They can be attached to reactive chemical groups compatible with chemical conjugation to proteins, and they are safe *in vivo*. Pegylated proteins have been approved for human use. Pegylated interferon is an example (Sharieff et al., *Cleve Clin J Med* 69:155, 2002). Pegylation effectively enhances the size of the resulting conjugate while avoiding immunogenicity or activity alterations. However, PEG has its own chemical and physical characteristics, and this can alter a conjugates ADME. For example, PEG alters the distribution of IL2 in such a way as to unacceptably increase its toxicity (Chen et al., *The Journal of Pharmacology and Experimental Therapeutics* 293:248, 2000). Also, the chemical conjugation is difficult to completely control, and any resulting conjugate is likely to be a mix of chemical species.
- 20 Another promising technique involves conjugating or fusing proteins to a carrier protein. There are many examples of chimeric molecules formed either through chemical reaction between the parent proteins or through the fusion of their gene sequences. In the case of fusion proteins, experience shows that the separate polypeptides constituting a

fusion protein generally fold into their three dimensional conformation independently. In fact, often a recombinant protein that misfolds during expression in *E. coli* by itself will fold properly when fused to a protein that regularly folds correctly. Examples include fusions to commercially available proteins such as GST and NusA (see for example

5 Novagen, Madison, WI).

One technique to make therapeutic fusion proteins is to fuse native therapeutics to human serum albumin (HSA) (U.S. Pat. No. 5,876,969). HSA is a 66 kD protein that is abundant in the human bloodstream. It is non-immunogenic and readily available. Potential problems include changed distribution of any resulting conjugate and the effect

10 of HSA as it is shuttled into cells that normally do not contain it intracellularly.

Another technique is to make therapeutic homomultimer fusion proteins. In this case, the coding DNA sequence for a functional protein is connected to copies of itself. A dimer of superoxide dismutase ("SOD") is disclosed in U.S. Pat. No. 5,084,390, whereby the hinge region of an immunoglobulin joins two copies of the SOD monomer.

15 The resulting dimer has an extended *in vivo* half-life. In another example, a dimer of erythropoietin is disclosed in U.S. Pat. No. 6,242,570.

Methods to manufacture highly polymerized sequences, for example polymers having greater than two units, have been developed in the field of artificial protein polymers. Lewis et al (*Protein Expression and Purification* 7:400, 1996) reveal a method 20 utilizing compatible, but nonregenerable, overhang restriction sites that are engineered to allow the polymerization of a monomeric spider silk repeating sequence in a geometric fashion. In similar manner, Elmorani, et al. (*Biochemical and Biophysical Research Communication* 239:240, 1997) use compatible, but nonregenerable, blunt end restriction sites to produce a polymeric form of wheat gliadin.

25 The techniques disclosed in both cases are predicated on the presence of a pair of compatible, nonregenerable, restriction sites at the end of the polymerizing protein sequence. This requirement severely limits the number of sequences that are amenable to polymerization. Another disadvantage of currently available methods is that once a final polymeric sequence is generated, the researchers must employ additional steps to 30 engineer it with the appropriate 5' and 3' sequences for expression.

SUMMARY OF THE INVENTION

- The present invention provides methods to easily and quickly generate multimers, such as dimers and higher order multimers, of DNA sequences and their open reading frame protein translations, resulting in constructs for the expression of proteins of greater molecular weight and valency. Methods are described whereby a sequence is attached to one or more versions of itself, either via a direct fusion or with a linker, where each version shares strong homology and is generally considered the same via its sequence and mode of action. In addition, the multimer is attached to terminal functional elements.
- 5 The monomer can theoretically have any sequence and can consist of elements from one or more genes or synthetic DNA fragments. Thus, although the polymerization employs homomultimers, the fundamental monomers themselves can be generated from heterogeneous sequences. Furthermore, heteromultimers can be produced from monomers previously manipulated with the methods of this invention if the constitutive monomers have compatible ends.
- 10 15

In one aspect, the present invention comprises multimer assemblies of cassettes that comprise nucleic acid sequences having restriction sites that can be ligated together to form constructs (multimer cassettes) having multiple copies of a sequence of interest (the monomer sequence), such as a sequence that encodes a peptide or protein.

- 20 Restriction sites used to ligate cassettes of a multimer assembly together to make a multimer cassette comprise restriction pair members that when ligated together, do not regenerate a restriction site. In one embodiment of the present invention, multimer assemblies are used that comprise 1) at least one amplification cassette comprising at least a monomer sequence and 2) at least one 3'-terminal cassette comprising at least one 3' specific sequence or at least one 5'-terminal cassette comprising at least one 5' specific sequence. Preferably, the 5'-terminal and/or 3'-terminal cassettes additionally comprise at least a portion of the monomer sequence.
- 25

- In some preferred embodiments of this aspect of the invention, component cassettes (such as amplification cassettes, 5'-terminal and/or 3'-terminal cassettes) of a multimer assembly can comprise one or more flanking restriction sites that can facilitate cloning of multimer cassettes.
- 30

In some preferred embodiments, component cassettes (such as amplification cassettes, 5'-terminal and/or 3'-terminal cassettes) can comprise one or more linker sequences, such as linker sequences that encode amino acids or peptides that can be used to link monomers. Such linker sequence can also comprise restriction sites, such as 5 restriction pair members that can be used in making multimer cassettes.

In another aspect, the present invention provides methods of making multimer cassettes. Such methods include ligation of 3' and 5' restriction pair members of component cassettes. In some preferred embodiments, the synthesis of multimer cassettes can optionally make use of flanking restriction sites that can be provided in the 10 component cassettes. In some preferred embodiments, the synthesis of multimer cassettes can optionally make use of restriction sites that can be provided in linker sequences included in one or more component cassettes.

The protein polymers encoded by DNA multimers of a multimer cassette can be expressed in any suitable gene/protein expression system. For example, prokaryotic or 15 eukaryotic systems are suitable, as are *in vitro* translation systems. The multimer assembly system described here facilitates the multimerization process and enables the production of multimers of any size and with a variety of N-terminal, linker, and C-terminal elements from a limited number of starting DNA sequences. For example, a gene can be designed for intracellular expression with an N-terminal methionine and for 20 extracellular expression by including a secretory signal sequence after the N-terminal methionine.

The invention can be used to produce constructs having multimeric or polymeric sequences of increased size and multiplicity.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagram showing an example of a multimer assembly and its cassettes for monomers having a terminal restriction pair. (A) shows a 5'-terminal cassette with sequence elements coding for protein N-terminal elements. The crosshatched elements are restriction sites, the rectangular segments are portions of the monomer sequence, the looping arrows indicate continuation as a plasmid, straight arrows indicate linker sequences, and ~ refers to arbitrary DNA sequences. The circle is a start codon, and the square is a 5' specific sequence. Restriction site 1 can include the start codon and/or can be a flanking restriction site for cloning flexibility. Restriction site 3 is the 3' restriction pair member, and 2 and 4 are flanking restriction sites for cloning flexibility. (B) shows an amplification cassette with sequence elements coding for a polymerizing sequence. Restriction site 5 is the 5' restriction pair member. (C) shows a 3'-terminal cassette with sequence elements coding for C-terminal elements. The pentagon represents 3' specific sequence and the hexagon a stop codon. The restriction site arrangement is preferred, but not the only arrangement for construction of an insert cassette. (D) shows one example of a Linker sequence. As shown here, it can contain elements 5' and 3' of the restriction pair formed by ligating restriction sites 5 and 3 together. The left and right arrows represent linker 5' and 3' elements, respectively.

Fig. 2 is a diagram showing one example of a multimer assembly and its cassettes for a monomer with an internal restriction pair. The crosshatched elements are restriction sites, the rectangular segments are portions of the monomer sequence, the looping arrows indicate continuation as a plasmid, straight arrows indicate linker sequences, and ~ refers to arbitrary DNA sequences. The circle is a start codon, and the square is a 5' specific sequence. The pentagon represents 3' specific sequence and the hexagon a stop codon. (A) shows a 5'-terminal cassette with sequence elements coding for N-terminal elements. (B) shows an amplification cassette with sequence elements coding for the polymerizing sequence. The double arrow represents a linker (optional). (C) shows a 3'-terminal cassette with sequence elements coding for C-terminal elements. (D) shows an

alternative 3'-terminal cassette that requires use of sequential ligation to form a multimer expression cassette.

Fig. 3 is a diagram showing two examples of pathways that can be used in the polymerization of amplification cassettes. Both procedures depicted involve two generalized cassettes, one with insert sequence b1 and the other with insert sequence b2. For pathway A, the b2 containing cassette is opened by digesting with enzymes 1 and 5. The b1 insert sequence is isolated after digesting the b1 containing cassette with enzymes 1 and 3. For pathway B, the b1 containing cassette is opened by digesting with enzymes 2 and 3. The b2 insert sequence is isolated after digesting the b2 containing cassette with enzymes 2 and 5. The final ligations to generate multimer assemblies are similar for both cases. The crosshatched elements are restriction sites, the rectangular segments are insert sequences, the looping arrows indicate continuation as a plasmid, and ~ refers to arbitrary DNA sequences.

15

Fig. 4 is a diagram showing examples of sequential ligation of cassettes to create a functional multimer cassette of a desired size. The schematic is a generalization of the sequential ligation procedure necessary for use with a 3'-terminal cassette given in Figure 2D. Pathway A depicts the insertion of an 'S' plasmid fragment into a 'T' containing plasmid, while Pathway B depicts the insertion of a 'T' plasmid fragment into a 'S' containing plasmid. In the figure, $S + T = 5I + AI$, $AI + 3I = 5IAI + 3I$, or $5I + AI3I = 5IAI3I$, where $5I \equiv$ the insert from a 5'-terminal cassette, $AI \equiv$ the insert from an amplification cassette, $3I \equiv$ the insert from a 3'-terminal cassette, $5IAI \equiv$ the insert resulting from the ligation of $5I$ and AI , $AI3I \equiv$ the insert resulting from the ligation of AI with $3I$, and $5IAI3I \equiv$ the insert resulting from the ligation of $5I$ with $AI3I$ or $5IAI$ with $3I$. Formation of $5IAI3I$ requires two sequential ligations and generation of intermediate $5IAI$ or $AI3I$ cassettes for each polymer size made. The crosshatched elements are restriction sites, the rectangular segments are insert sequences, the looping arrows indicate continuation as a plasmid, and ~ refers to arbitrary DNA sequences.

20
25
30

Fig. 5 is a diagram showing possible methods for generation of an insertion cassette. Pathways A and B are alternative pathways for insertion cassette generation based on different arrangements of flanking restriction sites. Pathway A involves opening the 5'-terminal cassette and inserting a fragment from the 3'-terminal cassette, while Pathway B involves opening the 3'-terminal cassette and inserting a fragment from the 5'-terminal cassette. The crosshatched elements are restriction sites, the rectangular segments are portions of the monomer sequence, the looping arrows indicate continuation as a plasmid, straight arrows indicate linker sequences, and ~ refers to arbitrary DNA sequences. The circle is a start codon, and the square is a 5' specific sequence. The pentagon represents 3' specific sequence and the hexagon a stop codon.

Fig. 6 is a diagram showing one possible method of generating a functional multimer cassette of a desired size from an insertion cassette and an amplification cassette. The insertion cassette is opened at both sites of the restriction pair with subsequent ligation of the insert from an amplification cassette, but the insert can ligate in the wrong orientation. Correct inserts must be identified by subsequent analysis. The crosshatched elements are restriction sites, the rectangular segments are portions of the monomer sequence, the looping arrows indicate continuation as a plasmid, straight arrows indicate linker sequences, and ~ refers to arbitrary DNA sequences. The circle is a start codon, and the square is a 5' specific sequence. The pentagon represents 3' specific sequence and the hexagon a stop codon.

Fig. 7 is a diagram showing another possible method of generating a functional multimer cassette of a desired size from an insertion cassette and an amplification cassette. The insertion cassette is opened with enzymes 3 and 2 to create an oriented ligation, but an additional step is required. In this case, the amplification cassette has flanking restriction site 2 on the 3' side of restriction site 3. The crosshatched elements are restriction sites, the rectangular segments are portions of the monomer sequence, the looping arrows indicate continuation as a plasmid, straight arrows indicate linker sequences, and ~ refers to arbitrary DNA sequences. The circle is a start codon, and the square is a 5' specific sequence. The pentagon represents 3' specific sequence and the hexagon a stop codon.

Fig. 8 is a diagram showing another possible scheme for generating a functional multimer cassette of a desired size from an insertion cassette and an amplification cassette in similar fashion to **Figure 7**, but the amplification cassette has flanking restriction site 2 5 on the 5' side of restriction site 5.

Fig. 9 is a diagram showing the PCR amplification of the hGH gene, its subsequent ligation to generate p0A0, and the ligation of the OmpA leader sequence to generate p0C0A2.

10 **Fig. 10** is a diagram showing the PCR mutagenesis of the hGH gene to generate p0A01. The diagram also shows the ligation of the OmpA sequence into p0A01 to generate p0A11A2 and the ligation of the PstI/BamHI fragment from p0A01 into P0A03 to generate p0A11A1.

15 **Fig. 11** is a diagram showing the PCR mutagenesis of the hGH gene to generate p0A11B.

Fig. 12 is a diagram showing the ligation of synthetic sequences to generate p0A11C1 and p0A11C2.

20 **Fig. 13** is diagram showing the polymerization of a GH direct fusion amplification cassette.

25 **Fig. 14** is diagram showing the generation of the GH direct fusion insertion cassette, p0A11D, and subsequent ligation of an amplification cassette to generate a multimer expression cassette.

Fig. 15 is a diagram showing the PCR mutagenesis of the hGH gene to generate p0A21B, the base amplification cassette for the GH glycine linker assembly.

30

Fig. 16 is a diagram showing the PCR mutagenesis of the hGH gene to generate the base cassettes, p0A31A, p0A31B, and p0A31C, for the GH SWG₄S assembly.

Fig. 17 is a diagram showing the sequential ligation of the GH SWG₄S assembly

5 cassettes to generate the multimer expression cassette, p0A31E3.

Fig. 18 is a picture of an SDS-PAGE gel showing the separation of proteins by molecular weight from separate lysates from cells expressing different polymers of rhGH. Lane 1 contains molecular weight standards, lane 2 the rhGH monomer, lane 3 the rhGH dimer, 10 lane 4 the rhGH trimer, lane 5 the rhGH pentamer, and lane 6 the rhGH nanamer.

Fig. 19 is a diagram showing insertion of synthetic sequences to generate the G₄S assembly 5'-terminal and amplification cassettes.

15 **Fig. 20** is a diagram showing PCR mutagenesis of the hGH gene to generate p0A04 and p0A41C.

Fig. 21 is a diagram showing ligation of the insert from p0D13A with p0A04 to generate p0A43B and ligation of the PstI/EcoRI fragment from p0A11A1 to generate p0A43A.

20

Fig. 22 is a diagram showing ligations to generate the base cassettes; p0A51A, p0A51B, and p0A51C, for the GH direct fusion assembly utilizing blunt ended HindIII and NcoI sites for the restriction pair.

25

Fig. 23 is a diagram showing the polymerization of the p0A51B insert to generate p0A51B2.

DETAILED DESCRIPTION OF THE INVENTION

Introduction

The current invention discloses methods that extend the polymerization techniques in three important ways. First, it introduces new methods to generate highly polymerized sequences from monomers that are incompatible with previous protein polymerization techniques. Second, it introduces additional linker sequences that, when paired with the monomer sequences, facilitate their use. Third, it introduces methods that facilitate the construction and expression of functional multimers and polymers. Taken together, the new methods enable the generation of large numbers of polymer variants that can differ in sequence and degree of polymerization. These variants can then be tested for desirable traits.

The disclosed techniques are applicable to any polypeptide sequence and can prove useful for proteins for which increased total molecular weight is deemed advantageous. The disclosed techniques are also useful for proteins for which increased valency is deemed advantageous. For example, expression of single chain antibody fragments fused together as larger multimers have the advantage of high valency and a stable linkage. Furthermore, if cassettes for two different sequences share compatible restriction pair members, they can be co-polymerized to produce heteromultimers.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Conventional methods are used for these procedures, such as those provided in the art and various general references. Where a term is provided in the singular, the inventors also contemplate the plural of that term. The nomenclature used herein and the laboratory procedures described below are those well known and commonly employed in the art. Where there are discrepancies in terms and definitions used in references that are incorporated by reference, the terms used in this invention shall have the definitions given herein. As employed throughout the disclosure, the following terms, unless otherwise indicated, shall be understood to have the following meanings:

Monomer. A DNA or amino acid sequence whose polymerization is desirable. A monomer can be a portion of a naturally occurring sequence (for example, a binding domain of an antibody). The sequence can be derived from one or more naturally occurring ones, or can be a synthetic sequence, or can be any combination of sequences of synthetic and natural origins. Monomers of the present invention can comprise linkers. As used herein monomer sequence means a nucleic acid sequence.

Multimer. A nucleic acid sequence encoding two or more monomers.

10 Polymer or Multimeric protein. A functional polypeptide that can be synthesized from a multimer assembly of the present invention. A polymer comprises at least two monomers (where each monomer can optionally comprise one or more linkers), can comprise one or more 5' translated regions (for example, signal peptides, N-terminal regions, "pro" or "pre" protein sequences, tag sequences, etc.), and can comprise one or more 3' translated 15 regions (for example, C-terminal regions, tag sequences, etc.)

20 Linker. A linker is a DNA or amino acid sequence that connects one DNA sequence with another through covalent bonds or an amino acid or peptide that connects one peptide or protein unit with another peptide or protein unit through peptide bonds. An amino acid or peptide linker can be a single amino acid (for example, glycine) or can be more than one amino acid.

25 Restriction Pair. Two restriction sites that have different recognition sequences that are ligation compatible, but when ligated together do not regenerate either of the two original restriction sites. A restriction pair can include two restriction sites that have overhangs, such as BglII and BamHI, or can include any two blunt end restriction sites that do not have the same recognition sequence, such as StuI and NaeI. In a broader application, a restriction pair can also include restriction sites that are initially ligation incompatible but are blunt ended to make them ligation compatible. An example includes blunt ending 30 HindIII and NcoI to make them ligation compatible.

Restriction pair member or restriction member. A restriction site that is part of a restriction pair. The 5' and 3' restriction pair members together make up a restriction pair, and each is the other's partner.

- 5 5' restriction pair member or 5' restriction member or 5' member. A restriction pair member that is located at the 5' terminus of a DNA sequence, such as a DNA sequence that, at least in part, encodes a monomer whose multimerization is desired or multimer of the present invention, or is located at the 5' terminus of a DNA sequence of interest whose ligation to a multimer is desired. The term "5' restriction pair member" or "5' member" can be used to refer to an unaltered restriction site (for example, a Bam HI site) or to a restriction site that has been altered, such as, for example, a filled-in 5' restriction pair member (such as blunt ended Bam HI site), or a fused 5' restriction pair member (for example, a ligated BamHI/BglII site).
- 10 3' restriction pair member or 3' restriction member or 3' member. A restriction pair member that is located at the 3' terminus of a DNA sequence, such as a DNA sequence that, at least in part, encodes a monomer whose multimerization is desired or multimer of the present invention, or is located at the 3' terminus of a DNA sequence of interest whose ligation to a multimer is desired. The term "3' restriction pair member" or "3' member" can be used to refer to an unaltered restriction site (for example, a BglII site) or to a restriction site that has been altered, such as, for example, a filled-in 3' restriction pair member (such as blunt ended BglII site), or a fused 3' restriction pair member (for example, a ligated BamHI/BglII site).
- 15 Flanking restriction site or flanking site. A restriction site that is not a member of a restriction pair used in the constructs and methods of the present invention. Its location outside of insert sequences and restriction pair members used in the cassettes and methods of the present invention can facilitate manipulation of the insert.

Insertion restriction site. A specific flanking restriction site that is 3' of the 3' restriction pair member of the 5'-terminal cassette and 5' of the 5' restriction pair member of the 3'-terminal cassette.

- 5 Amplification cassette. A DNA sequence that includes at least one monomer that is flanked by a restriction pair. An amplification cassette has a 5' restriction pair member at its 5' terminus and a 3' restriction pair member at its 3' terminus. The restriction pair enables the multimerization of the sequence or the ligation of it to other sequences with ligation compatible restriction sites. An amplification cassette can optionally comprise
10 other sequences as well, such as but not limited to sequences that code for amino acid or peptide linkers.

- 5' -terminal cassette. A DNA sequence that comprises a 3' restriction pair member, at least one 5'-specific sequence, where a 5'-specific sequence is a sequence that, when positioned at the 5' end of a multimer sequence, can facilitate the use of DNA multimers or the expression, purification, or identification of at least one protein polymer of the present invention, and, preferably, at least a portion of a monomer sequence. The 3' restriction pair member is ligation compatible with the 5' terminus of at least one amplification cassette. The 5' -terminal cassette is useful for introducing 5' -terminal
20 DNA sequences that contribute to making a sequence functional. Examples of 5' specific sequences include, but are not limited to, the translation start codon, secretion sequences, tag sequences, linker sequences, or special restriction sites.

- 3' -terminal cassette. A DNA sequence that comprises a 5' restriction pair member, at least one 3'-specific sequence, where a 3'-specific sequence is a sequence that, when positioned at the 3' end of a multimer sequence, can facilitate the use of DNA multimers or the expression, purification, or identification of at least one protein polymer of the present invention, and, preferably, at least a portion of a monomer sequence. The 5' restriction pair member is ligation compatible with the 3' terminus of at least one amplification cassette. The 3' -terminal cassette is useful for introducing 3' -terminal
30 DNA sequences that contribute to making a sequence functional. Examples of 3' specific

sequences include, but are not limited to, tag sequences, C-terminal sequences, polyadenylation sequences, stop codons, linker sequences, and the like.

Insert sequence. The functional sequence in a cassette. For the amplification cassette, 5 the functional sequence includes both restriction pair members and all sequence in between, including the monomer sequence. For the 5'-terminal cassette, the functional sequence includes the 3' restriction pair member, all 5'-specific sequences, and its portion of a monomer sequence, if present. For the 3'-terminal cassette, the functional sequence includes the 5' restriction pair member, all 3'-specific sequences, and its 10 portion of a monomer sequence, if present. For multimer cassettes, the functional sequence includes the functional sequences of the constitutive cassettes.

Multimer assembly. The collection of all cassettes that, in combination, after ligation, yields functional multimer DNA sequences or polymer protein sequences of a starting 15 monomer. A multimer assembly comprises one or more 5'-terminal cassettes and one or more amplification cassettes; one or more amplification cassettes and one or more 3'-terminal cassettes; or one or more 5'-terminal cassettes, one or more amplification cassettes, and one or more 3'-terminal cassettes that can be fused using 3' and 5' restriction pair members.

20

Multimer cassette. A cassette resulting from the ligation of two or more cassettes from the same multimer assembly.

Insertion Cassette. A multimer cassette generated from the ligation of a 5'-terminal and 25 3'-terminal cassette of a multimer assembly that is ligation compatible with any of said assembly's amplification cassettes to generate a multimer cassette.

Multimer expression cassette. A multimer cassette that, when transcribed and translated in a suitable expression system, produces a polymer protein sequence of a starting 30 monomer.

Segment of a monomer sequence. A segment of a monomer sequence is a portion of monomer sequence, that is, a nucleic acid sequence that encodes a portion of a monomer.

I. METHODS OF MAKING MULTIMER ASSEMBLIES

5

The present invention includes methods of fusing two or more nucleic acid sequences. The nucleic acid sequences can encode for peptide or protein sequences, such that when the nucleic acid sequences are expressed, a polymeric protein is produced. Preferably, in the methods of the present invention, the peptide or protein monomers encoded by the nucleic acid sequences are identical peptide or protein monomers. However, this is not a requirement of the present invention. The nucleic acid sequence, whose polymerization is desired is called a monomer sequence.

Monomer sequences can encode proteins or peptides whose function is known or unknown. Preferably, however, the identity and function of the peptide or protein encoded by a monomer sequence is known. Of particular interest are peptides and proteins that can have diagnostic or therapeutic value (for example, human growth hormone, hGH), although the invention is not limited to these protein sequences.

For example, monomer sequences can encode at least a portion of one or more receptors, receptor ligands, enzymes, inhibitors, transcription factors, translation factors, DNA replication factors, activators, chaperonins, or antibodies. Monomer sequences can also encode at least a portion of one or more cytokines, growth factors, or hormones such as, but not limited to, Interferon-alpha, Interferon-beta, Interferon-gamma, Interleukin-1, Interleukin-2, Interleukin-3, Interleukin-4, Interleukin-5, Interleukin-6, Interleukin-7, Interleukin-8, Interleukin-9, Interleukin-10, Interleukin-11, Interleukin-12, Interleukin-13, Interleukin-14, Interleukin-15, Interleukin-16, Erythropoietin, Colony-Stimulating Factor-1, Granulocyte Colony-stimulating Factor, Granulocyte-Macrophage Colony-Stimulating Factor, Leukemia Inhibitory Factor, Tumor Necrosis Factor, Lymphotoxin, Platelet-Derived Growth Factor, Fibroblast Growth Factors, Vascular Endothelial Cell Growth Factor, Epidermal Growth Factor, Transforming Growth Factor-beta, Transforming Growth Factor-alpha, Thrombopoietin, Stem Cell Factor, Oncostatin M, Amphiregulin, Mullerian-Inhibiting Substance, B-Cell Growth Factor, Macrophage

Migration Inhibiting Factor, Endostatin, and Angiostatin. Descriptions of these proteins can be found in Human Cytokines: Handbook for Basic and Clinical Research, Aggarwal, B. B. and Guterman, J. U. Eds., Blackwell Scientific Publications, Boston, Mass., (1992), which is herein incorporated by reference in its entirety.

5 The monomer encoding sequences are polymerized together by ligation of compatible, nonregenerable restriction sites, called restriction pair members. Unlike previous methodologies, the present invention employs cassettes with sequences other than those encoding the original monomer itself in the construction process. For example:

10 In the methods of the present invention, multimer assemblies are used that comprise at least one amplification cassette and at least one of the following: at least one 3'-terminal cassette or at least one 5'-terminal cassette. An amplification cassette comprises an insert sequence that includes a monomer sequence whose polymerization is desired, a 5' restriction pair member at its 5' terminus, and a 3' restriction pair member at 15 its 3' terminus. A 3'-terminal cassette comprises an insert sequence that includes at least one 3' specific sequence and a 5' restriction pair member site that can be fused to a 3' restriction pair member site of at least one of the one or more amplification cassettes. A 5'-terminal cassette, comprises an insert sequence that includes at least one 5' specific sequence and a 3' restriction pair member site that can be fused to a 5' restriction pair 20 member site of at least one of the one or more amplification cassettes. Preferably, the 5'-terminal and/or 3'-terminal cassettes additionally comprise at least a portion of the monomer sequence.

25 5' specific sequences can be, but are not limited to, sequences that enhance transcription, translation, secretion, protein folding, protein solubility, or binding of the protein to specific binding members such as antibodies. 3' specific sequences can be, but are not limited to, stop codons or sequences that enhance RNA stability, protein folding, protein solubility, or binding of the protein to specific binding members such as antibodies.

30 In the multimer assemblies of the present invention, 5' and 3' restriction pair members are used to fuse amplification cassettes, and preferably, where applicable, 3'-terminal cassettes to amplification cassettes and 5'-terminal cassettes to amplification

cassettes. 5' and 3' restriction pair members are preferably unique restriction sites that are ligation compatible, and said ligation destroys each member. In the alternative, 5' and 3' restriction pair members can be ligation incompatible sites that are made ligation compatible by blunt ending.

5 One aspect of the present invention is construction of cassettes comprising one or more flanking restriction sites that aid their use, but this is not a requirement of the present invention. Preferably, 3'-terminal cassettes and 5'-terminal cassettes, if present, comprise 3' and 5' flanking restriction sites. Flanking restriction sites can be any restriction site (except restriction pair member sites used in the same construct), and
10 preferably aid the use of cassettes by increasing the facility of making multimer cassettes. For example, the flanking sites facilitate the manipulation of the insert sequences, including their isolation and ligation. For example, some preferred methods employ an insertion restriction site, which is a specific flanking restriction site that is 3' of the 3' restriction pair member of the 5'-terminal cassette and 5' of the 5' restriction pair
15 member of the 3'-terminal cassette. Flanking restriction sites can also optionally be used to transfer constructs and assemblies to different expression vectors

In some preferred methods of the invention, sequences encoding linkers are employed. Multimer assembly cassettes can comprise one or more linker sequences. Multimer assembly cassettes can have linker sequences 5' of one or more insert
20 sequences, 3' of one or more insert sequences, or both 5' and 3' of one or more insert sequences. Linker sequences can be part of amplification cassettes, 5'-terminal cassettes, 3'-terminal cassettes, or any combination thereof. In preferred aspects of the present invention, nucleic acid sequences that encode amino acid or peptide linkers that are used to link monomers can also comprise restriction sites, such as 3' or 5' restriction pair
25 member sites that can facilitate construction of multimer assemblies. This provides a convenient means for introducing restriction pair members for efficient polymerization of monomer sequences through amplification cassettes and optionally 5'-terminal cassette or 3'-terminal cassette ligations. Alternatively, or in addition, amino acid or peptide linkers can be used to provide optimal spacing or folding of translated monomers or a
30 polymer.

Where more than one linker sequence is used in a single multimer assembly cassette, they may or may not occur between each and every monomer sequence. Where more than one linker sequence is used in a single multimer assembly cassette, they can encode the same or different amino acid or peptide linkers.

5 Peptide linkers are well known in the art. Preferably linkers are between one and twenty amino acids in length, and more preferably between one and ten amino acids in length, although length is not a limitation in the linkers of the present invention. Preferably linkers comprise amino acid sequences that do not interfere with the conformation and activity of peptides or proteins encoded by monomers of the present
10 invention. Some preferred linkers of the present invention are those that include the amino acid glycine. Examples include those disclosed in **Table 1**.

In an expressed protein polymer, such amino acid or peptide sequences join peptide or protein monomer sequences. If a linker is part of the insert sequence of the amplification cassette, it becomes part of the monomer that is to be multimerized. The
15 linker sequence can comprise at least one restriction pair member.

The present invention also introduces several methods to expand the use of restriction pair member sites. For example:

In some methods of the present invention, restriction pair members that are used to join monomer sequences are internal to a monomer sequence. In these embodiments, an
20 amplification cassette comprises a 5' segment of a monomer sequence and a 3' segment of a monomer sequence that together comprise the sequence of a complete monomer. The 5' segment is positioned 3' of the 3' segment, the 5' terminus of the 3' segment is a 5' restriction pair member, and the 3' terminus of the 5' segment is a 3' restriction pair member. In this case, in making a multimer cassette, ligation of the 3' restriction pair
25 member of the 5' segment of one amplification cassette with the 5' restriction pair member of the 3' segment of another amplification cassette can form a complete monomer sequence. In order to complete the polymer sequences, a multimer assembly preferably comprises a 5'-terminal cassette that comprises the 5' monomer segment and a 3'-terminal cassette that comprises the 3' monomer segment. In this way, monomer
30 sequences provided in the amplification cassettes can be provided in non-contiguous segments. In some preferred methods of the present invention, the amplification cassette

further comprises a linker that is positioned between the 5' segment and the 3' segment of the monomer sequence.

In some methods of the present invention, restriction pair members can be overhang restriction sites. In some methods of the present invention, restriction pair members can be blunt end restriction sites. In some other methods of the present invention, restriction pair members are incompatible "overhang" restriction sites that are converted to blunt end restriction sites through the use of polymerases or nucleases.

In some preferred methods of the present invention, restriction pair members are conveniently provided in one or more linker sequences. In these embodiments, linker sequences comprising a restriction pair member can be engineered onto the 3', 5', or both ends of an insert sequence.

In some preferred methods of the present invention, the 3'-restriction pair member codes for a stop codon that is destroyed upon ligation to the 5'-restriction pair member.

15 In one aspect of the present invention, the assembly methodology consists of the following four steps:

1. Generate or obtain the DNA for the monomer.

Techniques familiar to those skilled in the art include, but are not limited to:

- a. Amplification of a sequence from a DNA library, optionally including any additions or mutations to the sequence in PCR primers.
- b. Chemical synthesis of the sequence
- c. Splicing of sequences together from pre-existing DNA

20 2. Decide what linker sequence, if any, to use between monomers and construct a multimer assembly.

25 Options for the linker include none (direct fusion of monomers), a linker encompassing a restriction pair member within its sequence, a linker with restriction pair members at one or more termini, or a linker lacking a restriction pair member.

Once a linker is added, it becomes part of the monomer sequence.

30 For each option, three basic cassettes can be generated: one or more 5'-terminal cassettes, at least one amplification cassette, and one or more 3'-terminal cassettes.

However, in some instances, all three cassettes are not required. A multimer assembly comprises at least one amplification cassette, and one or more 5'-terminal cassettes or one or more 3'-terminal cassettes, or can have at least one amplification cassette, one or more 5'-terminal cassettes, and one or more 3'-terminal cassettes. In some cases, 5 multiple versions of each cassette may be desirable. Furthermore, the amplification cassette can be polymerized to produce new higher order (multimeric) amplification cassettes.

The ends of the monomers determine the characteristics of the cassettes. The current invention discloses the use of linkers to introduce ends containing a restriction pair as 10 well the construction of 5'-terminal and/or 3'-terminal cassettes to facilitate their use.

As an alternative to engineering the ends of a monomer with a restriction pair, then the cassettes can be constructed with a restriction pair internal to the monomer sequence. The construction of the cassettes is modified to accommodate the presence of an noncontiguous monomer in each.

15 Finally, a method is disclosed in which the constructions for a restriction pair either at the ends or internal to the monomer is extended to use with a pair of incompatible restriction sites. This method is less preferred, as the method requires that blunt ends for ligation are created for each ligation step (by nuclease digestion or polymerase fill-in, or both), decreasing the efficiency of the procedure.

20 The following are the general steps for construction of the assemblies for each possible restriction pair case:

a. Using a monomer sequence with a terminal restriction pair.

The scheme shown in **Figure 1** is applicable for any monomer sequence that can be engineered with a terminal restriction pair. The steps to engineer the assembly can 25 include the following:

(1) Engineer 5'-terminal cassettes containing one or more 5' specific DNA sequences (for example, start codon, secretion sequence, etc.), preferably the monomer sequence, linker sequence, if present, and the 3' member of the restriction pair.

(2) Engineer an amplification cassette containing a 5' restriction member, optionally a 30 first linker sequence, at least one monomer sequence, optionally a second linker sequence, and a 3' restriction member.

(3) Engineer 3'-terminal cassettes containing a 5' restriction member, optionally a linker sequence, preferably the monomer sequence, and one or more 3'-terminal specific DNA sequences (specific recognition sequences, stop codon, etc.).

5 An alternative formulation involves 5'-terminal and/or 3'-terminal cassettes that do not include any monomer sequence. The utility of including the monomer sequence in both terminal cassettes lies in utilizing the restriction pair members to join each terminal cassette to an amplification cassette, however, this is not a requirement of the present invention.

10

b. Using a monomer sequence with an internal restriction pair.

The scheme shown in **Figure 2** is applicable for any monomer sequence that can be engineered with an internal restriction pair. The steps to engineer the assembly include the following:

15

(1) Engineer 5'-terminal cassettes containing one or more 5' specific DNA sequences (start codon, secretion sequence, etc.), the portion of a monomer sequence that occurs on the 5' side of the restriction pair (the 5' monomer segment), and finally the 3' restriction pair member.

20

(2) Engineer an amplification cassette containing a 5' restriction pair member, DNA encoding the portion of a monomer sequence that occurs 3' of the restriction pair (the 3' monomer segment), optionally a linker sequence, DNA encoding the portion of a monomer that occurs 5' of the restriction pair (the 5' monomer segment), and a 3' restriction pair member.

25

(3) Engineer 3'-terminal cassettes containing the 5' restriction pair member, the portion of a monomer sequence that occurs 3' of the restriction pair (the 3' monomer segment), and one or more 3'-terminal specific DNA sequences (specific recognition sequences, stop codon, etc.).

30

c. Using a monomer sequence with a pair of incompatible restriction sites made compatible by blunt ending.

Either scheme shown in **Figure 1** or **Figure 2** are applicable, but in this case the restriction pair consists of restriction sites that are blunt ended to make them compatible.

Once constructed, the amplification cassette enables generation of a sequence containing any number of monomers fused together.

5

3. Polymerize the amplification cassette in an arithmetic, geometric, or mixed progression (see **Figure 3**).

A series of amplification cassettes are generated from the original amplification cassette. The technique involves digesting a first construct comprising an amplification cassette at two 5' or two 3' sites of an insert, one of which is a restriction pair member site and the other of which is an external flanking site (external to the restriction pair member site), to open up the construct. This is followed by digesting a second construct comprising an amplification cassette at the same flanking site, but with the opposite restriction pair member, to release the amplification sequence from the plasmid as a fragment. This sequence is then ligated into the opened first plasmid construct from before. Both restriction sites used in the ligation are destroyed, but the resulting cassette has intact flanking restriction sites and an intact restriction pair on the ends that enable further polymerizations.

Mixing and matching the cassettes used to open a construct that comprises an amplification cassette and to generate an insert from a construct that comprises an amplification cassette enables new cassettes of any size to be made in an arithmetic, geometric, or mixed progression. For example, if the monomer is used to both open the plasmid and create insert, a dimer cassette is made. If the resulting dimer is used for both, then a tetramer is made. If this tetramer is used for both, then an octamer is made, and continuation leads to a binomial geometric progression. On the other hand, if the monomer is always used as the insert and the newest cassette is used to receive the insert, an arithmetic progression of one is produced. For instance, when a dimer construct is opened and a monomer fragment inserted, then a trimer is produced. When a trimer construct is opened and a monomer fragment is inserted, then a tetramer is produced. In general, any new cassette can be mixed with any previously generated cassette to allow rapid generation of a polymer of any desired size. For example, if a polymer of size 20 is

desired, the 16mer is generated geometrically, and ligating the 16mer to the tetramer generates the 20mer in a total of only 5 ligation.

Subsequent ligation to 5'- and 3'-terminal cassettes can enable production of a functional multimer. The multimer's size, based on actual molecular weight, is approximately a whole number multiple of the original. In addition, the composition of the multimer is almost identical to the monomer, differing only because of any linker sequences or terminal flanking regions that are used.

It is important to note that the polymerization does not require flanking sites. Without flanking sites, the ligations can occur with the fragments joined in either orientation, and more laborious subsequent analysis is needed to identify the correct constructs. In contrast, use of flanking sites facilitates the process by enabling oriented ligations.

4. Ligate the cassettes together to give a full length, functional, multimer.

The cassettes can be ligated sequentially as shown in **Figure 4**, or an insertion cassette can be created from the 5'- and 3'-terminal cassettes as diagramed in **Figure 5** with subsequent insertion of the polymerized amplification cassette as shown in **Figures 6, 7, and 8**. The use of an insertion cassette expedites the creation of a series of multimers with the same 5' and 3' terminal elements. **Figure 6** illustrates a technique for the ligation of the fragment from an amplification cassette into an insertion cassette using only the restriction pair restriction sites. However, the ligation is not oriented, necessitating additional analysis to identify correct constructs. **Figures 7 and 8** show equivalent oriented ligations that result from different arrangements of flanking sequences.

Figure 4 illustrates a method of making a multimer cassette from two cassettes from a multimer assembly utilizing flanking sites comprising a first cassette comprising either a 5'-restriction pair member or a 3'-restriction pair member and a second cassette comprising both a 5'-restriction pair member and a 3'-restriction pair member and further comprising:

- 1) providing the first cassette with a first flanking restriction site at one end, either 5' or 3', of its insert sequence;
- 2) providing the second cassette with a second flanking restriction site that is, or is made, ligation compatible with the first flanking site and is on the same side, either 5' or 3', of its insert sequence as the first flanking restriction site is relative to the first cassette's insert sequence;
- 3) digesting the first cassette at its restriction pair member and the first flanking site and isolating the first fragment containing the insert sequence;
- 4) digesting the second cassette at its restriction pair member partner to the first cassette's restriction pair member and at the second flanking site and isolating the second fragment containing the insert sequence;
- 5) ligating the first fragment with the second fragment to generate a multimer cassette.

15

The identities of the first and second cassettes can vary. For example, the first cassette can be a 3'-terminal cassette and the second cassette an amplification cassette, the first cassette can be a 5'-terminal cassette and the second cassette an amplification cassette, the first cassette can be a 3'-terminal cassette and the second cassette a multimer cassette constructed from a 5'-terminal cassette and an amplification cassette, or the first cassette can be a 5'-terminal cassette and the second cassette a multimer cassette constructed from a 3'-terminal cassette and an amplification cassette.

20 For the case when the first cassette is a 3'-terminal cassette and the second cassette is an amplification cassette, if the amplification cassette is digested at its 3' restriction pair member and a flanking restriction site on the 5' side of its 5' restriction member to generate a ligatable fragment, then the 3'-terminal cassette is digested at its 5' restriction pair member and a flanking restriction site on the 5' side of this member to generate a ligatable cassette. Alternatively, if the amplification cassette is digested at its 3' restriction pair member and a flanking restriction site on the 3' side of this member to generate a ligatable cassette, then the 3'-terminal cassette is digested at its 5' restriction

pair member and a flanking restriction site on the 3' side of its complete insert to generate a ligatable fragment.

It is important to note that the ligation of cassettes together does not require flanking sites. However, flanking sites enable oriented ligations. For example, if flanking sites are absent, a method of making a multimer cassette from two cassettes from a multimer assembly comprising a first cassette comprising either a 5'-restriction pair member or a 3'-restriction pair member and a second cassette comprising both a 5'-restriction pair member and a 3'-restriction pair member comprises:

10. 1) digesting the first cassette at its restriction pair member and isolating the first fragment containing the insert sequence;
- 2) digesting the second cassette at both its restriction pair member sites and isolating the second fragment containing the insert sequence;
- 3) ligating the first fragment with the second fragment and screening for correct ligation orientation to generate a multimer cassette.

Again, the identities of the first and second cassettes can vary. The first cassette can be a 3'-terminal cassette and the second cassette an amplification cassette, the first cassette can be a 5'-terminal cassette and the second cassette an amplification cassette, the first cassette can be a 3'-terminal cassette and the second cassette a multimer cassette constructed from a 5'-terminal cassette and an amplification cassette, or the first cassette can be a 5'-terminal cassette and the second cassette a multimer cassette constructed from a 3'-terminal cassette and an amplification cassette.

Figure 5 illustrates a method of making an insertion cassette from the 5'-terminal cassette and the 3'-terminal cassette when each shares an insertion restriction site. The method comprises:

30. 1) providing the 5'-terminal cassette with a first flanking restriction site, independent of the insertion restriction site, that is outside of the sequence including the insert sequence and insertion restriction site of the 5'-terminal cassette;

- 2) providing the 3'-terminal cassette with a second flanking restriction site, independent of the insertion restriction site, that is outside of the sequence including the insert sequence and insertion restriction site of the 3'-terminal cassette and is , or is made, ligation compatible with the first flanking site and is on the same side, either 5' or 3', of its insert sequence as the first flanking restriction site is relative to the 5'-terminal cassette's insert sequence;
- 5
3) digesting the 5'-terminal cassette at its insertion restriction site and the first flanking site and isolating the first fragment containing the insert sequence;
- 10
4) digesting the 3'-terminal cassette at its insertion restriction site and the second flanking site and isolating the second fragment containing the insert sequence;
- 15
5) ligating the first fragment with the second fragment to generate an insertion cassette.

Figure 6 illustrates a method of making a multimer cassette comprising an insertion cassette and an amplification cassette from a multimer assembly comprising:

- 20
1) digesting the insertion cassette at both its restriction pair member sites and isolating the first fragment containing the insert sequence;
- 2) digesting the amplification cassette at both its restriction pair member sites and isolating the second fragment containing the insert sequence;
- 25
3) ligating the first fragment with the second fragment and screening for correct ligation orientation to generate a multimer cassette.

Figures 7 and 8 illustrate a method of making a multimer cassette comprising an insertion cassette and an amplification cassette comprising:

- 30
1) digesting the amplification cassette at the insertion restriction site and its restriction pair member on the opposite side, either 5' or 3', of the

- insert sequence and isolating the first fragment containing the insert sequence;
- 5 2) digesting the insertion cassette at the insertion restriction site and the restriction pair member partner to the digested amplification cassette's restriction pair member and isolating the second fragment containing the insert sequence;
- 10 3) ligating the first fragment with the second fragment to generate a multimer cassette precursor;
- 4) digesting the multimer cassette precursor at both restriction pair members, isolating the fragment containing the insert sequence, and ligating it with itself to generate a multimer cassette.

Once constructed, the gene for the multimer can be used as an insert to construct other cassettes or to express it in a suitable transcription and translation system. Once 15 isolated in the correct conformation and with the necessary degree of purity, polymeric polypeptides are available for applications in the fields of medicine, veterinary care, research and development, diagnostics, etc. The present invention comprises proteins made from multimer assemblies of the present invention.

Each cassette can involve a fusion of any of a number of functional elements. For 20 example, any construction involving a linker is by nature a heteromultimer, because the monomer contains at least two functional elements. A particularly expeditious method to produce these fusions is to treat each functional element as a nested assembly. In other words, each element itself is an assembly that consists of individual cassettes.

The current methods are easily extended to heteromultimers if two sequences 25 share compatible restriction sites. For instance, two distinct monomer amplification cassettes, A and B, can be ligated together if they share the same restriction pair. Subsequent polymerization of this new "monomer" results in an alternating sequence, ABAB... Any pattern of alternating sequences can theoretically be constructed from any number of initial monomers. For example, the pattern ABCABC... is just one 30 possibility.

II MULTIMER ASSEMBLIES AND MULTIMER CASSETTES

The present invention includes multimer assemblies made using the methods of the present invention and novel cassettes incorporating novel restriction pair members. In some preferred aspects of the present invention, a multimer assembly of the present invention comprises two or more amplification cassettes, in which fused 5' and 3' restriction pair member sites join the amplification cassettes. An amplification cassette can comprise any practical number of monomer sequences.

Multimer assemblies of the present invention comprise component constructs having 5' restriction pair members, 3' restriction pair members, or both 5' restriction pair members and 3' restriction pair members that can be used to make multimer cassettes, including multimer expression cassettes. Such cassettes are synthesized by joining component cassettes (such as 5'-terminal cassettes, 3'-terminal cassettes, and amplification cassettes) by ligating a 3' restriction pair member site of one component cassette to a 5' restriction pair member site of another component cassette.

One multimer assembly of the present invention comprises one or more amplification cassettes and at least one 3'-terminal cassette. Another multimer assembly of the present invention comprises one or more amplification cassettes and at least one 5'-terminal cassette. Another multimer assembly of the invention comprises one or more amplification cassettes, at least one 3'-terminal cassette, and at least one 5'-terminal cassette.

Multimer expression cassettes made from multimer assemblies of the present invention include, for example, multimer cassettes in which a 5'-terminal cassette is fused to an amplification cassette comprising a single monomer, multimer cassettes in which a 5'-terminal cassette is fused to a multimer amplification cassette constructed from multiple amplification cassettes, and multimer cassettes in which a 5'-terminal cassette is fused to a multimer cassette comprising one or more amplification cassettes and at least one 3'-terminal cassette. Multimer expression cassettes made from multimer assemblies of the present invention also include, for example, multimer cassettes in which a 3'-terminal cassette is fused to an amplification cassette, multimer cassettes in which a 3'-terminal cassette is fused to a multimer amplification cassette constructed

from multiple amplification cassettes, and multimer cassettes in which a 3'-terminal cassette is fused to a multimer cassette comprising one or more amplification cassettes and at least one 5'-terminal cassette.

The present invention also includes novel amplification cassettes. In one aspect of
5 the present invention, an amplification cassette comprises at least one linker, in which at least one of the one or more linkers comprises at least one restriction pair partner. Amplification cassettes can be fused using restriction pair partners, at least one of which is introduced in the linker, to form a multimer amplification cassette. The method of making the multimer amplification cassette is by joining two or more amplification
10 cassettes by ligating the first restriction pair partner of at least one of the two or more amplification cassettes to the second restriction pair partner of at least one other of the two or more amplification cassettes to generate a multimer cassette. The present invention includes multimer amplification cassettes comprising component amplification cassettes that incorporate linkers, and multimer assemblies and multimer expression
15 cassettes that include such multimer amplification cassettes.

Also included as amplification cassettes of the present invention are amplification cassettes that comprise monomer sequences in noncontiguous orientation. For example, an amplification cassette can comprise a 5' segment of a monomer sequence and a 3'
20 segment of a monomer sequence that together comprise the sequence of a complete monomer, in which the 5' segment is positioned 3' of the 3' monomer segment. In these embodiments, the 5' terminus of the 3' monomer segment is preferably a 5' restriction pair member and the 3' terminus of the 5' monomer segment is preferably a 3' restriction pair member. The present invention also includes multimer amplification cassettes
25 comprising two or more amplification cassettes that comprise monomer sequence in noncontiguous orientation. Such multimer cassettes comprising multiple amplification cassettes can be made by ligating a 3' restriction member of at least one of the two or more amplification cassettes to a 5' restriction member of at least one other of the two or more amplification cassettes. The present invention also includes multimer assemblies
30 and multimer expression cassettes that include such amplification and multimer amplification cassettes.

In yet another aspect, the present invention includes amplification cassettes that comprise 3' and 5' restriction pair members comprising restriction sites that are initially ligation incompatible but are blunt ended to make them ligation compatible. The present invention also includes multimer amplification cassettes comprising two or more amplification cassettes that comprise noncompatible sites that have been blunt-ended and then ligated to join the two or more amplification cassettes. The present invention also includes multimer assemblies and multimer expression cassettes that include such amplification and multimer amplification cassettes.

10 The invention includes multimer assembly cassettes in vectors, including cloning and expression vectors, where expression vectors can be designed for *in vitro* or *in vivo* expression. The vectors can be designed for *in vivo* expression in prokaryotes or eukaryotes, including but not limited to, bacterial cells, fungal cells, algal cells, plant cells, insect cells, avian cells, and mammalian cells. The present invention also 15 encompasses cells that include such vectors and polymeric proteins made using vectors that comprise multimeric expression vectors of the present invention. The present invention also encompasses polymeric proteins expressed from the multimeric assemblies of the present invention.

The disclosed invention also encompasses the construction of different multimer assemblies involving multimeric hGH, and multimer cassettes made using the methods of the present invention that comprise multimerized hGH sequences or multimerized portions of hGH. Sequences encoding hGH or portions thereof that are part of multimer cassettes and multimer assemblies of the present invention include sequences that encode hGH taking into account the redundancy of the genetic code. Sequences encoding hGH or 25 portions thereof that are part of multimer cassettes and multimer assemblies of the present invention include sequences that encode hGH can also comprise sequence changes with respect to the human GH gene sequence that change the amino acid sequence where such changes do not detrimentally affect the activity of the protein or portion thereof.

30 The hGH assemblies can differ in the functional elements included, such as those provided by 3' - or 5' -terminal elements. The ease of producing these assemblies, and the

resulting multimers and polymers, demonstrates the utility of the methods disclosed. In the examples below, restriction sites outside, and flanking, the restriction pair sites are engineered in order to facilitate the manipulation of the cassettes.

Endogenous hGH appears in several forms *in vivo* as a result of expression from more than one gene, as well as alternative gene splicing. The predominant mature form of hGH is a single polypeptide chain consisting of 191 amino acids. The DNA and protein sequences for this predominant form are given as SEQ ID NO: 1 and SEQ ID NO: 2, respectively.

In the following paragraphs, the term "engineer" refers to using standard techniques of molecular biology generally known to those skilled in the art. Standard techniques include, but are not restricted to, restriction digestion and ligation, PCR amplification and mutagenesis, DNA synthesis, DNA isolation and purification, etc., as described in Sambrook et al. (2000), which are hereby incorporated by reference. As such, the details are only described if they bear directly on the present invention or deviate from common practice.

Examples

A drawback to rhGH therapy is the need for once daily injections. Understandably, patient preference is for a minimum of injections. In an attempt to overcome this, rhGH has been formulated with PLGA in microspheres, chemically linked to PEG, and fused to HSA in order to produce longer acting versions. Here we describe the construction of families of multimeric rhGHs, according to the steps below using the general procedures shown in **Figures 1 to 8**.

25

Example 1

The first example involves isolation of the GH gene. Steps to isolate the hGH gene are summarized in **Figure 9**. hGH is highly expressed in the anterior pituitary gland. As a result, mRNA of hGH is abundantly found in lysates of human pituitary. The gene for hGH is PCR amplified from human pituitary cDNA (Human Pituitary Gland Quick-CloneTM cDNA, BD Biosciences Clontech, Palo Alto, CA, catalog #7173-

1) using SEQ ID NO: 3 as the 5' primer and SEQ ID NO: 4 as the 3' primer. The 5' primer has an NdeI restriction enzyme site coding for an N-terminal methionine, and the 3' primer has a BamHI restriction enzyme site immediately after the TAG stop codon. The resulting PCR fragment is isolated from the reaction mix using standard techniques,
5 as are all subsequent ones.

The purified PCR fragment is ligated into parent plasmid pET41a (Novagen, Madison, WI) after both insert and plasmid are digested with NdeI and BamHI and purified, again using standard techniques. This plasmid ligation mixture, and all others unless otherwise indicated, is transformed into DH5 α cells and plated on LB/antibiotic
10 plates. Single colonies are sub-cultured and plasmid DNA is isolated from each. Restriction enzyme analysis is used to confirm the presence of an insert into the plasmid, and plasmids with insert are sent for DNA sequencing using SEQ ID NO: 5 and SEQ ID NO: 6 (Novagen, Madison, WI) as amplification primers for the 5' and 3' ends,
15 respectively. Plasmid with correct insert is identified as p0A0, and the DNA coding region and corresponding open reading frame (ORF) translation are listed in SEQ ID NO:
7 and SEQ ID NO: 8, respectively. The convention for the sequences is that the restriction sites are included at the termini of DNA sequences and only translated amino acids that eventually appear in an expressed insert are given. Expression of protein from p0A0 yields a 192 amino acid protein consisting of full length hGH with an additional N-
20 terminal methionine.

It is convenient to engineer a high copy number plasmid that contains the hGH gene and enables digestion of the hGH gene in its interior so that 5' or 3' elements can be swapped in and out. The gene for hGH contains a convenient PstI site, CTGCAG. The plasmid p04 (SEQ ID NO: 9), a derivative of pUC19 (New England Biolabs) containing
25 the same multi-cloning site as pET41a, is first readied by digesting with PstI, followed by Mung Bean Nuclease, and subsequent re-ligation to destroy the internal PstI site to create p04A1. Finally, the NdeI/BamHI hGH fragment from p0A0 is ligated into similarly digested p04A1 to yield p0A03.

Several examples are now given to generate assemblies for GH multimers with
30 different linkers. Variation in the linker sequence, as well as the degree of monomer

polymerization, may alter the polymers ease of production, conformation, *in vitro* activity, *in vivo* activity, immunogenicity, etc.

Example 2

- 5 The second example involves generation of an assembly for the direct fusion multimer of GH.

There is not a convenient restriction pair at the termini of rhGH, so this example uses the methods for a monomer sequence with an internal restriction pair. A direct fusion assembly for hGH is constructed with the features diagrammed in **Figure 2**.

- 10 Disclosed are two 5'-terminal cassettes, the amplification cassettes, and two 3'-terminal cassettes. The 3'-terminal cassette is engineered to enable construction of an insertion cassette, as shown in **Figure 5**. This facilitates insertion of amplification cassettes to generate expressible genes for different size homopolymeric GHS.

- Two 5'-terminal cassettes for the GH fusion protein assembly are disclosed. The
15 first is a direct start 5'-terminal cassette, and the second is an OmpA start 5'-terminal cassette. The direct start results in an N-terminal methionine at the N-terminus of the final expressed GH polymer. Its construction is straight forward because the insert in p0A0 and p0A03 already has the N-terminal methionine fused to the GH gene. In contrast, the OmpA start codes for an N-terminal leader sequence that targets the polymer
20 to the periplasmic space of *E. coli*, resulting in the cleavage of the leader from the polymer. There are many other 5'-terminal cassettes that can easily be generated by those skilled in the art.

- A pre- 5'-terminal cassette is disclosed that enables fusion of the OmpA sequence to any other blunt end or HindIII digested sequence. SEQ ID NO: 10 is a synthetic DNA
25 fragment that contains the coding sequence for the OmpA leader peptide, and its ORF translation is listed in SEQ ID NO: 11. The fragment has a 5' NdeI site, the OmpA leader coding region, a 3' HindIII site for HindIII ligation or blunt end ligation after filling in the HindIII 5' overhang with T4 DNA polymerase, and a BamHI site for cloning flexibility. Plasmid p04 is readied by digestion to destroy an internal site, this
30 time the HindIII site. The plasmid is digested with HindIII, followed by Mung Bean Nuclease, and subsequently ligated back together to create p04A2. Both p04A2 and

insert DNA are digested with NdeI and BamHI and ligated together to yield the plasmid p0C0A2 as shown in **Figure 9**.

For the current use, a GH sequence is needed that contains a 5' blunt end or HindIII site, along with a 3' restriction site that is the 3' member of a restriction pair.

- 5 The 5' terminus is engineered with a HindIII site. Digestion with Mung Bean Nuclease after digestion with HindIII results in a blunt 5' end that leaves the 5'-terminal codon of GH, TTC, intact. Although the blunt end is not needed for the current example, in general it is necessary for ligation to other hypothetical cassettes.

There are several choices for the restriction site pair, and we choose to use GH 10 amino acids 187 and 188, glycine and serine, that are compatible with, among other enzymes, BamHI and BclI. The two enzymes recognize sequences GGATCC and TGATCA, respectively. BamHI is assigned as the 3' member, and BclI is assigned as the 5' member.

The desired DNA sequence is generated by PCR using p0A03 as template, as 15 shown in **Figure 10**. The 5' and 3' primers are listed in SEQ ID NO: 12 and SEQ ID NO: 13, respectively, and the DNA coding region for the insert between the 5' flanking NdeI and 3' BamHI sites is listed in SEQ ID NO: 14. The fragment is digested with HindIII and BamHI and inserted into similarly digested p04B1 to yield p0A01. Plasmid p04B1 is prepared by destroying the HindIII site in p04A1 as described for the 20 preparation of p04A2. The result is a parent plasmid with the PstI and HindIII sites destroyed.

The 5'-terminal cassettes are now constructed from the generated sequences as 25 shown in **Figure 10**. The XbaI/HindIII fragment from p0C0A2 is inserted into plasmid p0A01 to generate p0A11A2. The result is the OmpA 5'-terminal cassette for the GH direct fusion assembly. It contains the OmpA sequence fused directly to the 5' coding region of GH. The resulting DNA insert between NdeI and BamHI is listed in SEQ ID NO: 15, with corresponding ORF listed in SEQ ID NO: 16. The direct translation start 5'-terminal cassette is constructed by ligating fragments from existing sequences. The PstI/BamHI 5' GH fragment and plasmid backbone that results from digesting p0A03 is 30 ligated with the PstI/BamHI 3' GH fragment that results from digesting p0A01 to yield p0A11A1. The resulting DNA sequence between NdeI and BamHI, and the

corresponding ORF, for p0A11A1 are listed in SEQ ID NO: 17 and SEQ ID NO: 18, respectively.

As shown in **Figure 2**, the amplification cassette must contain several components. First, it must have both the 5' and 3' members of the restriction pair to enable polymerization. In between must be the entire continuous GH sequence. Finally, if convenient, there should be flanking restriction sites for insertion and extraction of the sequence from a plasmid backbone.

The amplification cassette for the current direct fusion of GH is generated by PCR, as shown in **Figure 11**. The 5' primer is listed in SEQ ID NO: 19. It contains an NdeI site, the 5' restriction pair member BclII, followed by the codons that together code for GH amino acids 187-191, and finally codons to anneal to the GH 5'-terminal codons. The 3' primer is one previously used and listed in SEQ ID NO: 13. The PCR template is p0A03. The resulting insert DNA sequence between NdeI and BamHI is listed in SEQ ID NO: 20, with ORF sequence listed in SEQ ID NO: 21. The DNA sequence is inserted into plasmid p04A1 to yield p0A11B.

Two simple 3'-terminal cassettes are disclosed, as shown in **Figure 12**. Both code for the 3' terminus of GH, starting at the glycine and serine codons within the BclII site, amino acids 187 and 188, and ending with the translation stop codon, TAG. The first cassette, given in SEQ ID NO. 22, is a direct translation stop. The double stranded DNA is synthesized and contains an EcoRI site flanking the 5' terminus, a BclII site to ligate to BamHI, the 3' terminus of GH, a stop codon, and a SalI site for cloning flexibility. It is inserted into p04A1 by digesting the synthetic DNA and p04A1 with EcoRI and SalI and ligating the large fragments together to yield plasmid p0A11C1. The C-terminal ORF protein sequence contributed by this cassette to subsequent GH multimer constructs is given in SEQ ID NO: 23.

The second 3'-terminal cassette, given in SEQ ID NO: 24, is a synthetic DNA fragment similar to the first, except it contains the codons for a 3 amino acid polylysine tail before the stop codon. It is analogously inserted into p04A1 to yield plasmid p0A11C2. The polylysine tail is potentially useful for chemical conjugation with other molecules. SEQ ID NO: 25 is the C-terminal ORF sequence contributed by the new insert to subsequent GH multimer constructs.

Once the basic cassettes are complete, the amplification cassette can be polymerized, the 5'-terminal and 3'-terminal cassettes can be joined to form an insertion cassette, and finally amplification cassettes can be ligated to the insertion cassette to generate expressible multimers.

5

Example 3

The polymerization of the GH direct fusion amplification cassettes is performed as shown in general in **Figure 3** and specifically in **Figure 13**. The first polymerization is formation of the dimer. Plasmid p0A11B is digested with NdeI and BclI and the 10 plasmid isolated. In a separate reaction, p0A11B is digested with NdeI/BamHI and the insert isolated. The two fragments are then ligated together to yield plasmid p0A11B2. Its insert DNA sequence is listed in SEQ ID NO: 26, and the corresponding ORF translation is listed in SEQ ID NO: 27. This process is repeated, changing the identity, and thus the size, of amplification cassettes 1 and 2 in **Figure 13** to construct polymer 15 inserts of different sizes. The size of new constructs is increased fastest if the polymerization is done geometrically, each time using the most recent construct for both cassettes 1 and 2. The size is increased by one if the monomer amplification cassette, p0A11B, is used either as cassette 1 or 2. The generalized sequences for the resulting 20 amplification cassettes are given in SEQ ID NO: 28 and SEQ ID NO: 29 for the DNA and protein, respectively.

Example 4

The cassettes for the GH direct fusion assembly are designed to enable construction of insertion cassettes to facilitate generation of a variety of expressible polymers. The 25 general procedures are shown in **Figures 5** and **7** and the specifics in **Figure 14**. Different insertion cassettes can be generated with the various 5'-terminal and 3'-terminal cassettes. However, only the one involving p0A11A1 and p0A11C1 is described here. Others are constructed in exactly the same way.

Plasmid p0A11A1 is digested with EcoRI and SalI and the opened plasmid is 30 isolated. Plasmid p0A11C1 is digested with the same enzyme pair and the insert isolated. The two fragments are ligated together to generate the insertion cassette, p0A11D, and

the resulting DNA sequence is listed in SEQ ID NO: 30. Plasmid p0A11D is compatible with ligation of any of the amplification cassettes for this assembly. It need be prepared only once for all subsequent ligations, as long as the supply is sufficient.

5

Example 5

Either of the two schemes shown in **Figures 6 and 7** can be used to ligate amplification cassettes into the insertion cassette. The example given here utilizes the oriented ligation shown in **Figure 7** and subsequent digestion and re-ligation to generate final products as shown in **Figure 14**.

10 Plasmid p0A11D is digested with BamHI and EcoRI, and the plasmid is isolated. An amplification cassette is digested with BclI and EcoRI and the insert isolated. Ligation of the two fragments yields an intermediate that is converted to the multimer expression cassette after digestion with BamHI and BclI, purification, and subsequent re-ligation. The result is an expression ready insert for the direct fusion growth hormone
15 multimer. When performed with the Nmer amplification cassette, the result is an N+1 multimer expression cassette. The insert has general DNA sequence listed in SEQ ID NO: 31 and corresponding ORF translation listed in SEQ ID NO: 32. The production of different size multimers is controlled by the size of the ligated amplification cassette.

20 Protein expression is achieved by digesting and ligating the multimer expression cassette insert into an appropriate expression system. For example, the insert can be liberated with NdeI and SalI and ligated into similarly digested pET41a, followed by transformation into *E. coli* strain BL21(DE3) (Novagen).

25 One utility of the invention is the ease of production of different size multimers and different variations once the basic cassettes, p0A11A1, p0A11A2, p0A11B, p0A11C1, and p0A11C2, for example, are constructed. Those skilled in the art can easily see how substituting p0A11C2 for p0A11C1 when generating the insertion cassette generates a polylysine tail variant.

Example 6

30 The next example involves generation of a GH multimer with a linker without a convenient restriction pair. The one amino acid linker, glycine, is used as an example.

The construction of GH multimers with a glycine linker is analogous to the construction of the fusion protein. In fact, the GH glycine linker assembly shares the same 5'- and 3'-terminal cassettes with the GH fusion protein assembly. This is one advantage of the assembly construction scheme given in **Figure 2**. Assemblies differing only in the linker
5 region only need different amplification cassettes, while sharing the same 5'- and 3'-terminal cassettes.

Use p0A11A1 and p0A11A2 as before for the direct start and OmpA 5'-terminal cassettes for the direct fusion assembly. Use p0A11C1 and p0A11C2 as before for direct stop and poly lysine 3'-terminal cassettes.

10 The only difference is the amplification cassettes that contain a glycine codon between the ending and starting codons for GH. The glycine linker amplification cassette is made in the same way as the one for the direct fusion homomultimer except for some necessary substitutions of sequences, as shown in **Figure 15**. SEQ ID NO: 33 is substituted for SEQ ID NO: 19 as the 5' PCR primer. It contains the same elements as
15 before, as well as the glycine codon between the sequence for amino acids 191 and 1. The resulting PCR fragment is inserted into parent plasmid p04A1 by digesting both the parent plasmid and the PCR fragment with NdeI and BamHI and ligating the appropriate fragments together. The resulting plasmid is labeled p0A21B. The DNA sequence and ORF translation for the insert sequence between NdeI and BamHI are listed in SEQ ID
20 NO: 34 and SEQ ID NO: 35, respectively.

25 The construction of additional amplification assemblies, the insertion cassette, and multimer expression cassettes for the GH glycine linker assembly is identical in practice to the one for the GH direct fusion assembly, **Figures 13 and 14**, except for the substitution of p0A21B for p0A11B. The corresponding generalized amplification cassette insert DNA and ORF sequences are listed in SEQ ID NO: 36 and SEQ ID NO:
37, and the general formulas for the multimer expression cassettes are listed in SEQ ID NO: 38 and SEQ ID NO: 39.

30 The previous examples have demonstrated, among other things, the ease at which multiple 5'- and 3'-terminal cassettes can be used to introduce variations in the N- and C-termini of a polymer. In the case of the 5'-terminal cassettes, cassettes with either a direct translation start or one introducing a leader sequence are disclosed. In the case of

the 3'-terminal cassettes, ones with either a direct stop or one introducing a polylysine tail are disclosed. Each demonstrates the ease at which functional elements can be added to the beginning or end of a polymer sequence. These methods are easily extended to other examples by those skilled in the art. Therefore, subsequent examples will be
5 limited to the presentation of only a single 5'- and 3'-terminal cassette for each assembly.

The next examples involve generation of GH multimers utilizing linkers that result in monomers with a terminal restriction pair. **Figure 1** details the general features for these assemblies.

10 Example 7

This example involves a linker that is noteworthy because it contains a 3' restriction pair member with a functional stop codon that is destroyed upon polymerization. Use of this linker makes it possible to express functional multimers using just the 5'-terminal and amplification cassettes. However, a 3'-terminal cassette is
15 necessary to express homomultimers without any residual linker at the 3' terminus of the protein.

The 5' restriction pair member is NcoI, C^CATGG, while the 3' restriction pair member is RcaI, T^CATGA. Therefore, the resulting linker sequence is A-Ser-Trp-B, where A and B are arbitrary protein sequences. For the given example, A is a null
20 sequence, and B is G₄S, where the single letter amino acid abbreviations are used.

For this example, only one 5'-terminal cassette is disclosed, with a direct ATG start codon and no leader sequence, as shown in **Figure 16**. The PCR primers for the 5'-terminal cassette are listed in SEQ ID NO: 3 and SEQ ID NO: 40, for the 5' and 3' ends, respectively. The 5' primer maintains the NdeI site and its start codon, while the 3'
25 primer introduces a stop codon within an RcaI (or BspHI) restriction site, immediately followed by a BamHI site. The template for the reaction is p0A0.

Because the RcaI restriction site also contains the codon TCA immediately 5' of the stop codon, it also introduces a C-terminal serine residue. The resulting PCR fragment is purified and ligated into pET41a in an analogous manner for the generation
30 of p0A0. The sequence verified plasmid is labeled p0A31A, and the DNA coding region, from the NdeI to the BamHI site, and the resulting ORF protein sequence are listed in

SEQ ID NO: 41 and SEQ ID NO: 42, respectively. Expression of protein from the gene for p0A31A yields a 193 amino acid protein consisting of full length hGH with an additional N-terminal methionine and C-terminal serine.

The PCR primers for the amplification cassette are listed in SEQ ID NO: 43 and
5 SEQ ID NO: 40, for the 5' and 3' ends, respectively. The 5' primer introduces an NcoI site followed by the linker region. The NcoI site is ligation compatible with the 3' RcaI site, and any such ligation destroys the TGA stop codon by altering it to a TGG codon. The resulting PCR fragment is purified and ligated into pET41a after the PCR product and plasmid are cut with NcoI and BamHI, as shown in **Figure 16**. The sequence
10 verified plasmid is labeled p0A31B, and the DNA coding region from the NcoI to the BamHI site is listed in SEQ ID NO: 44. The ORF protein sequence coded by the insert is given in SEQ ID NO: 45.

Again, for this example, only one 3'-terminal cassette is disclosed, with a direct TAG stop codon and no other 3'-specific sequences. The 3'-terminal cassette is
15 constructed using PCR with p0A0 as template and SEQ ID NO: 43 and SEQ ID NO: 4 as 5' and 3' primers, respectively. This creates a cassette with a 5' linker and a 3' stop codon immediately following the last amino acid from the parent monomer. The PCR fragment is inserted into pET41a as before and shown in **Figure 16** to create p0A31C. The resulting DNA and protein fragments between the NdeI and BamHI sites are listed in
20 SEQ ID NO: 46 and SEQ ID NO: 47, respectively.

The scheme for the polymerization of the amplification cassettes is shown in **Figure 3**. Additional care is necessary because the parent plasmid contains RcaI sites. One way to unambiguously liberate the insert sequence for polymerization is to first digest the flanking BamHI site, isolate the insert, and then digest with RcaI. The general
25 formulas for the Nmer amplification cassette are listed in SEQ ID NO: 48 and SEQ ID NO: 49 for the DNA and corresponding ORF translation, respectively.

Example 8

The ligation of the multimer assembly cassettes must be done sequentially, as
30 shown in **Figure 4**, because the arrangement of the restriction sites in the 3'-terminal cassette is like **Figure 2d**. The first ligation involves the 5'-terminal and amplification

cassettes, rather than the 3'-terminal and amplification cassettes, to take advantage of the stop codon in the 3'-restriction member to produce expression ready inserts. The specifics are shown in **Figure 17** using procedures already described. Use of the monomeric amplification cassette, p0A31B, results in the dimeric cassette, p0A31F2,

- 5 with insert DNA and corresponding ORF translation listed in SEQ ID NO: 50 and SEQ ID NO: 51. The general formulas for the N+1mer produced after ligation between the Nmer amplification and the 5'-terminal cassettes are listed in SEQ ID NO: 52 and SEQ ID NO: 53. Transfer of the insert into an appropriate expression system yields expression of the N+1 GH polymer with the SWG4S linker and C-terminal S residue.

10 Completion of the ligation scheme shown in **Figure 17** results in an insert with an additional monomer and the natural C-terminus of GH. If the insert from p0A31F2 is ligated into p0A31C, then the trimer expression cassette p0A31E3 is generated. In general, the formulas for the insert DNA and corresponding ORF translation when the Nmer amplification cassette is used are listed in SEQ ID NO: 54 and SEQ ID NO: 55.

- 15 For p0A31E3, the monomer amplification cassette is used and N=1.

Example 9

The plasmids containing the inserts generated with the ligation scheme shown in **Figure 17** are capable of expressing rhGH polymers following standard techniques (see for example, user manuals from Novagen, Madison, WI). DNA sequences listed in SEQ ID NO: 52 with N = 0, 1, 2, 4, and 8 and prepared according to Example 8 are ligated into pET41a. The resulting plasmids are separately transformed into BL21(DE3) and separately grown in Luria Broth medium and induced to express the polymer protein by adding IPTG to a concentration of 1 mM.

25 Following 3 hours of induction, each culture is harvested by centrifugation and treated with SDS-PAGE sample buffer. Proteins from the samples for each culture are separated according to their molecular weights on a standard SDS-PAGE gel (Invitrogen, Carlsbad, CA). The resulting gel is stained with coomasie blue stain to visualize the protein bands. Results for the monomer (SEQ ID NO: 42), dimer (N=1 in SEQ ID NO: 53), trimer (N=2 in SEQ ID NO: 53), pentamer (N=4 in SEQ ID NO: 53), and nanamer

(N=8 in SEQ ID NO: 53) are given in **Figure 18**. As the figure demonstrates, large amounts of each polymeric rhGH are produced except for the nanamer.

Example 10

5 Linkers with convenient restriction sites offer the engineering option to generate a multitude of assemblies with cassettes that can be attached to monomers using restriction/ligation techniques. The utility of this formulation lies in the breadth of assemblies that can be constructed relatively easily. This is especially apparent when the linkers themselves are treated as assemblies nested within the construction of the
10 multimers. Once constructed, these linker assemblies and cassettes, like any other, can be reused to produce new assemblies.

Nested linker assemblies are constructed having a slightly different function than the multimer assemblies. They still need an amplification cassette for the polymerization of the linker. However, the other cassettes in the assembly enable attachment of the
15 linker to either a 5' or 3' terminus, whichever is appropriate.

The example given here is a series of linkers, having amino acid sequence GZGS, where Z is an arbitrary sequence of arbitrary length. The series of linkers in **Table 1** below share features that enable them to be treated similarly in terms of their engineering. All but one has a Glycine at the N-terminus of the linker that can be coded by an NaeI
20 restriction site at the 5' end for blunt end ligation of a 5'-terminal cassette to a monomer pre-cassette. For the other linker, GS, a synthetic DNA fragment must be ligated to the monomer pre-cassette without propagation within a plasmid. Each of the linkers ends in the protein sequence GS, so that the restriction pair is identical to earlier examples utilizing the BclI and BamHI sites.

25

Table 1

Linker protein monomer unit	5'-terminal cassette DNA sequence	Amplification cassette DNA sequence
GS	GGATCC	TGATCAGGATCC
GGS	GCCGGCGGATCC	TGATCAGGCGGATCC
GGGS	GCCGGCGGCGGATCC	TGATCAGGCGGCGGATCC
GGGGS	GCCGGCGGCGGCGGATCC	TGATCAGGCGGCGGCGGATCC
GZGS	GCCGGCYGGATCC	TGATCAGGCYGGATCC
	Z is an arbitrary protein sequence, and Y is its DNA coding sequence.	

5

As a single example of the engineering of the linker assembly, we construct the $(G_4S)_x$ linker, where x indicates the degree of polymerization of the monomer sequence. The assembly is engineered like any other, and it falls into the scheme shown in **Figure 1**. The specifics are shown in **Figure 19**.

10

Two synthetic DNA sequences are needed, SEQ ID NO: 56 and SEQ ID NO: 57. The first, the 5'-terminal cassette labeled as p0D11A in **Figure 19**, is the sequence enabling addition of the linker sequence to other cassettes. It is flanked by a NcoI site, and thus with an upstream NdeI site, for cloning flexibility at the 5' terminus, contains the NaeI site to create the blunt end ligation with the glycine codon at the 5' terminus, the linker sequence, and finally the BamHI site within the GS codons. Plasmid p04 is prepared by digestion with NgoMIV, digestion with Mung Bean Nuclease, and finally re-ligation to destroy the internal NaeI site, creating plasmid p04A3. This altered plasmid, along with the insert, is digested with NcoI and BamHI and the appropriate fragments are ligated together. The resulting plasmid is labeled p0D11A. The open reading frame translation between the cleaved NaeI and the entire BamHI sites is G₄S.

15

20

SEQ ID NO: 57 is the sequence for the amplification cassette to create multimers of the G₄S linker. It is flanked by an NcoI site, again for cloning flexibility. It has the 5' BclI site from the restriction pair, followed by the G₄S coding sequence that ends with the BamHI site. It is inserted into p04 by cutting both plasmid and insert with NcoI and 5 BamHI and ligating the appropriate fragments together, as shown in **Figure 19**. The resulting plasmid is labeled p0D11B.

Amplification cassette p0D11B is polymerized by the scheme shown in **Figure 3**, left hand side, to create a dimer. In this instance the decision to follow the left hand side scheme results in larger fragments that are easier to isolate. Plasmid p0D11B is digested 10 with NdeI and BclI and the large fragment is isolated. Separately, the same parent plasmid is digested with NdeI and BamHI, this time isolating the small fragment. The two isolated fragments are then ligated together, destroying the internal BclI and BamHI sites, but preserving the flanking ones. The resulting plasmid is labeled p0D11B2, the DNA insert is listed in SEQ ID NO: 58, and the ORF translation is listed in SEQ ID NO: 15 59. The sequence codes for the dimer (G₄S)₂. The process can be repeated with different starting cassettes to generate any (G₄S)_x linker. In this manner, (G₄S)₄ can be generated by digesting p0D12B with NdeI and BclI and saving the large fragment and ligating in the small fragment generated by digesting it with NdeI and BamHI.

The engineering of the G₄S assembly enables the construction of a GH multimer 20 assembly with the (G₄S)₃ linker. The (G₄S)₃ 5'-terminal cassette for ligation to the GH sequences is generated following the general scheme shown in **Figure 4**. Plasmid p0D11B2 is digested with NdeI and BclI, and the large fragment is isolated. The small fragment resulting from digestion of p0D11A with NdeI and BamHI is ligated in, creating plasmid p0D13A. The DNA and ORF sequences for the insert are listed in SEQ 25 ID NO: 60 and SEQ ID NO: 61, respectively. The insert in p0D13A enables ligation of the (G₄S)₃ linker to the 3' end of any sequence ending in a blunt end.

Example 11

Engineering of the GH (G₄S)₃ assembly requires two new ends to the GH gene. 30 The BclI 5' restriction pair member is needed on the 5' terminus of the amplification and 3'-terminal cassettes, and a blunt end immediately after the last codon of GH is needed

on the 3' terminus of the 5'-terminal and amplification cassettes for ligation of the (G₄S)₃ linker. There are many ways to get a blunt end at the 3' terminus of GH. Disclosed here is the use of an NcoI site that is made blunt after digestion with Mung Bean Nuclease. In addition, it is convenient to introduce a stop codon flanked by the SalI restriction site at 5 the 3' terminus of the GH gene for construction of an insertion cassette, as shown in general in **Figure 5**.

Three new primers are used to generate the new termini on two new GH inserts by PCR using P0A03 as template, as shown in **Figure 20**. The 5' primer is listed in SEQ ID NO: 62. It contains a flanking NdeI site, the BclI 5' restriction pair member, and 10 sequence complementary to the GH 5' terminus. It is used for both PCR reactions. The 3' primers are listed in SEQ ID NO: 63 and SEQ ID NO: 64. Both contain sequence complementary to the 3' terminus of GH. The first codes for the NcoI site at the 3' terminus for creation of a blunt end after the last GH base pair and a flanking EcoRI site, while the second introduces a stop codon followed by a SalI restriction site.

15 The PCR fragments are ligated into plasmid backbones as shown in **Figure 20**. The PCR fragment resulting from use of the primers listed in SEQ ID NO: 62 and SEQ ID NO: 63 is digested with NdeI and EcoRI and ligated into similarly digested p04A1 to yield p0A04, while the fragment resulting from use of the primers listed in SEQ ID NO: 62 and SEQ ID NO: 64 is digested with BclI and SalI and ligated into similarly digested 20 p0A11C1 to give p0A41C. The insert in p0A04 between the BclI and blunt ended NcoI sites has the DNA sequence listed in SEQ ID NO: 65 and corresponding ORF translation listed in SEQ ID NO: 66. Likewise, the insert in p0A41C, the 3'-terminal cassette, between the BclI and SalI sites has the DNA sequence listed in SEQ ID NO: 67 and ORF translation listed in SEQ ID NO: 68.

25 The amplification cassette is generated first by ligating the (G₄S)₃ linker from plasmid p0D13A with the insert in p0A04, as shown in **Figure 21**. Plasmid p0D13A is digested with NaeI and HindIII, and the small fragment is isolated. It is ligated into p0A04 after digestion first with NcoI, then Mung Bean Nuclease, and finally HindIII to yield p0A43B. The resulting DNA sequence for the amplification cassette between the 30 BclI and BamHI sites is listed in SEQ ID NO: 69, with corresponding ORF translation in SEQ ID NO: 70.

The direct start 5'-terminal cassette is generated by combining the 5' elements from p0A11A1 with the 3' elements from p0A43B, as shown in **Figure 21**. The small fragment resulting from digesting p0A43B with PstI and EcoRI is isolated. It is ligated to the large fragment resulting from digestion of p0A11A1 with the same enzymes to yield 5 p0A43A. The DNA sequence for the insert between NdeI and BamHI is listed in SEQ ID NO: 71, with corresponding ORF translation in SEQ ID NO: 72.

The polymerization of the amplification cassettes again follows the scheme in **Figure 3**. The general formulas for the insert DNA and corresponding ORF translation for the Nmer amplification cassette are listed in SEQ ID NO: 73 and SEQ ID NO: 74.

10 The ligation of the cassettes for the GH (G₄S)₃ linker assembly to create a multimer expression cassette follows the previously described scheme shown in **Figure 7** and demonstrated in Example 4. The insertion cassette is first generated with the 5'- and 3'-terminal cassettes using EcoRI and SalI digestions. An amplification cassette insert is first isolated after digestion with BclI and EcoRI and then spliced into the insertion 15 cassette after digestion using BamHI and EcoRI. The resultant construct is subsequently digested with BamHI and BclI and re-ligated. The resulting N+2 multimer expression cassette, where N is the degree of polymerization of the amplification cassette used, has DNA and corresponding ORF translation sequences listed in SEQ ID NO: 75 and SEQ ID NO: 76. Transfer of the insert into a suitable expression system yields multimeric GH 20 with (G₄S)₃ linker.

Example 12

The last example is an alternative construction for a GH direct fusion assembly. It involves the use of an incompatible restriction pair that is blunt ended for ligation.

25 Construction of this new assembly is done by ligating together fragments from earlier cassettes, since they already contain the needed elements. The construction scheme is shown in **Figure 22**.

The 5'-terminal cassette is labeled p0A51A. It is generated by combining elements from p0A11A1 and p0A04. Plasmid p0A11A1 is digested with PstI and EcoRI 30 and the open plasmid isolated. This is ligated with the insert isolated after digesting

p0A04 with the same enzymes. The result, p0A51A, has DNA and corresponding ORF translation listed in SEQ ID NO: 77 and SEQ ID NO: 78.

The amplification and 3'-terminal cassettes are constructed in exactly the same manner as the 5'-terminal cassette, except for substituting which plasmids are digested.

- 5 For the amplification cassette, plasmid p0A01 is ligated with the insert from p0A04. The insert DNA and corresponding ORF sequences are listed in SEQ ID NO: 79 and SEQ ID NO: 80. Likewise, for the 3'-terminal cassette, plasmid p0A01 is ligated with the insert from p0A03. Its insert DNA and corresponding ORF translation are listed in SEQ ID NO: 81 and SEQ ID NO: 82.

- 10 The polymerization of amplification cassettes still follows the scheme in **Figure 3**. However, digestion at a restriction pair member now requires the additional blunt ending of its overhang. **Figure 23** shows the specifics for the current assembly. The digestions of the cassette are done sequentially so that the restriction pair is blunt ended, but the flanking restriction sites are left intact. The general formulas for the amplification
15 cassettes are listed in SEQ ID NO: 83 and SEQ ID NO: 84.

- The ligation of the multimer assembly cassettes is done sequentially as shown in **Figure 4**. The digestion of any plasmid is performed as described above with blunt ending of the restriction pair member first. The general formulas for the resulting multimer expression cassette insert, using the Nmer amplification cassette, are listed in
20 SEQ ID NO: 85 and SEQ ID NO: 86.

In practice, ligations of cassettes from this assembly involves more steps, but the technique's almost universal applicability may make it the method of choice in some instances. For the current case, the assembly given in Examples 1-4 is easier to manipulate.

* * * * *

Those skilled in the art will recognize many equivalents to the examples presented herein, using different monomers, linkers, restriction pairs, flanking restriction sites, 5' 5 specific sequences, 3' specific sequences, and ligation strategies. For example, the methods are flexible as to the order of ligating 5'-terminal cassettes, 3'-terminal cassettes, and amplification cassettes, and in ligating amplification cassettes to one another to form higher order amplification cassettes. Combining elements of the following claims 10 presented here and in the description, including the examples, is within the scope of the invention and are encompassed in the following claims.

All references cited herein, including the bibliography, are incorporated by reference in their entireties.

References Cited

U.S. PATENT DOCUMENTS

5

5,084,390	1/1992 Hallewell et al.	435/188
5,876,969	3/1999 Fleer et al.	435/69.7
6,242,570	6/2001 Sytkowski	530/350

10

OTHER PUBLICATIONS

Jorgensen, J.O.L., "Human Growth Hormone Replacement Therapy: Pharmacological and Clinical Aspects," *Endocrine Reviews*, 12(3):189, (1991).

15

Chien, Y.W., "Human Insulin: Basic Sciences to Therapeutic Uses," *Drug Development and Industrial Pharmacy*, 22(8):753-789 (1996).

20

Putney, S.D., and P. Burke, "Improving Protein Therapeutics with Sustained-release Formulations," *Nature Biotechnology*, 16:153, (1998).

Johnson, O.L., et al., "A Month-long Effect from a Single Injection of Microencapsulated Human Growth Hormone," *Nature Medicine*, 2(7):795, (1996).

25

Chen, S.A., et al., "Plasma and Lymph Pharmacokinetics of Recombinant Human Interleukin-2 and Polyethylene Glycol-modified Interleukin-2 in Pigs," *The Journal of Pharmacology and Experimental Therapeutics*, 293(1):248, (2000).

30

Venkatachalam, M.A., and H. Rennke, "The Structural and Molecular Basis of Glomerular Filtration," *Circulation Research*, 43(3):337-347, (1978).

- Roberts, M.J., et al., "Chemistry for Peptide and Protein PEGylation," *Adv Drug Deliv Rev*, 54(14):459-76, (2002).
- Sharieff, K.A., et al., "Advances in Treatment of Chronic Hepatitis C: 'Pegylated' Interferons," *Cleve Clin J Med*, 69(2):155-9, (2002).
- Lewis, R.V., et al., "Expression and Purification of a Spider Silk Protein: A New Strategy for Producing Repetitive Proteins," *Protein Expression and Purification*, 7:400-6, (1996).
- 10 Elmorjani, K., et al., "Synthetic Genes Specifying Periodic Polymers Modeled on the Repetitive Domain of Wheat Gliadins: Conception and Expression," *Biochemical and Biophysical Research Communication*, 239:240-6, (1997).
- 15 Pennell, C.A., and P. Eldin, "In vitro Production of Recombinant Antibody Fragments in *Pichia pastoris*," *Res Immunol*, 149(6):599-603, (1998).
- Sambrook, et al., "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, NY, (2000).
- 20 Aggarwal, B. B. and Guterman, J. U. Eds., "Human Cytokines: Handbook for Basic and Clinical Research," Blackwell Scientific Publications, Boston, Mass., (1992).

SEQUENCE LISTING

5 <110> Gentide Biopharmaceuticals, Inc.

Bussell, Stuart

10

<120> METHODS TO CONSTRUCT MULTIMERIC DNA AND POLYMERIC PROTEIN
SEQUENCES AS DIRECT FUSIONS OR WITH LINKERS

15

<130> GNT-00101.P.1-US

20

<150> US 60/396,466

<151> 2002-07-16

25

<160> 86

30

<170> PatentIn version 3.0

35

<210> 1

<211> 573

<212> DNA

40

<213> Homo sapiens

45

<400> 1
ttcccaacca ttcccttatac caggctttt gacaacgcta tgctccgcgc ccatcgcttg
60

caccagctgg cctttgacac ctaccaggag tttgaagaag cctatatccc aaaggaacag
120

50

aagtattcat tcctgcagaa cccccagacc tccctctgtt tctcagagtc tattccgaca
180

55

ccctccaaca gggagggaaac acaacagaaa tccaacctag agctgctccg catctccctg
240

ctgctcatcc agtcgtggct ggagcccgta cagttcctca ggagtgtt cgccaacagc
300

5 ctggtgtacg ggcctctga cagcaacgtc tatgacctcc taaaggacct agaggaaggc
360

atccaaacgc tgatggggag gctggaagat ggcagcccc ggactgggca gatcttcaag
420

10 cagacctaca gcaagttcga cacaaactca cacaacgatg acgcactact caagaactac
480

gggctgtct actgcttcag gaaggacatg gacaaggctcg agacattctt ggcacatcg
540

15 cagtccgct ctgtggaggg cagctgtggc ttc
573

20 <210> 2

<211> 191

<212> PRT

25 <213> Homo sapiens

30 <220>

<221> mat_peptide

35 <222> (1)...()

40 <400> 2

Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg
1 5 10 15

Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu
45 20 25 30

Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro
35 40 45

50 Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg
50 55 60

Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu
65 70 75 80

55 Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val

5 <210> 5
10 <211> 19
5 <212> DNA
10 <213> Artificial

15 <220>
15 <223> synthetic sequence
15 <400> 5
15 taatacgaact cactatagg
19

20 <210> 6
20 <211> 21
25 <212> DNA
25 <213> Artificial

30 <220>
30 <223> synthetic sequence
35 <400> 6
35 tgcttagttat tgctcagcgg t
21

40 <210> 7
40 <211> 588
45 <212> DNA
45 <213> Artificial

50 <220>
50 <223> synthetic sequence
55 <400> 7
55 catatgttcc caaccattcc cttatccagg cttttgaca acgctatgct ccgcgcggcat
60

cgtctgcacc agctggcctt tgacacctac caggagttt g aagaagccta tatcccaaag
120

5 gaacagaagt attcattcct gcagaacccc cagacctccc tctgtttctc agagtctatt
180

ccgacaccct ccaacaggga ggaaacacaa cagaaatcca accttagagct gctccgcata
240

10 tccctgtgc tcatccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgcc
300

aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
360

15 gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccccggac tgggcagatc
420

20 ttcaagcaga cctacagcaa gttcgacaca aactcacaca acgtgacgc actactcaag
480

aactacgggc tgcgtctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540

25 atcgtgcagt gccgctctgt ggagggcagc tgtggcttct agggatcc
588

30 <210> 8

<211> 192

<212> PRT

35 <213> Artificial

40 <220>

<223> synthetic sequence

<220>

45 <221> mat_peptide

<222> (1)..()

50 <400> 8

Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
1 5 10 15

55 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
20 25 30

Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 35 40 45

5 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
 50 55 60

Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
 65 70 75 80

10 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
 85 90 95

Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
 15 100 105 110

Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
 115 120 125

20 Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
 130 135 140

Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
 145 150 155 160

25 Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
 165 170 175

Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
 30 180 185 190

<210> 9

35 <211> 2907

<212> DNA

<213> Artificial

40 <220>

45 <223> synthetic sequence

<400> 9
 ccggatatacg ttccctcctt cagcaaaaaa cccctcaaga cccgtttaga ggccccagg
 60

50 gtttatgcta gttattgctc agcggtggca gcagccaact cagtttcctt tcgggctttg
 120

ttttagcagcc taggtattaa tcaatttagtg gtgggtgggg tgggtgggt gctcgagtgc
 180

55 ggccgcaagc ttgtcgacgg agctcgccctg caggcgcc aaggcctgta cagaattcgg
 240

atccccata tcccatggga ctcttgcgt cgtcatcacc ggagccacca ccggtaccca
300

5 gatctggct gtccatgtgc tggcgttcga atttagcagc agcggttct ttcataccaa
360

ttgcagtaact accgcgtggc accagacccg cgaggatgatg gtgatggta tgaccagaac
420

10 cactagtaca cacatatgta tatctcccttc ttaaagttaa acaaaattat ttcttagaggg
480

gaattgttat ccgctcacaa ttcccata gtgagtcgtat ttaatttcgc gggatcgaga
540

tcgatctcga tcctctacgc cggacgcatac gtggccggca tcaccggcgc cacaggtgcg
600

20 gttgctggcg cctatatcgc cgacatcacc gatgggaaag atcgggctcg ccacttcggg
660

ctcatgagcg cttgtttcgg cgtgggtatg gtggcaggcc ccgtggccgg gggactgttg
720

25 ggcgccatct cttgcatgc atggcgtaat catggtcata gctgtttcct gtgtgaaatt
780

gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt aaagcctggg
30 840

gtgcctaattt agttagctaa ctcacattaa ttgcgttgcg ctcactgccc gctttccagt
900

35 cggaaacct gtcgtgccag ctgcattaaat gaatcgccca acgcgcgggg agaggcggtt
960

tgcgtattgg gcgcctttcc gcttcctcgc tcactgactc gctgcgtcg gtcgttcggc
1020

40 tgcggcgagc ggtatcagct cactcaaagg cggttaatacg gttatccaca gaatcagggg
1080

ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg
45 1140

ccgcgttgct ggcgttttc cataggctcc gccccctga cgagcatcac aaaaatcgac
1200

50 gctcaagtca gaggtggcga aacccgacag gactataaag ataccaggcg tttccccctg
1260

gaagctccct cgtgcgtct cctgttccga ccctgcccgt tacggatac ctgtccgcct
1320

55 ttctcccttc gggaaagcgtg gcgctttctc atagctcacg ctgttaggtat ctcagttcgg
1380

tgttaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct
1440

5 ggcgccttatac cggttaactat cgtcttgagt ccaacccgggt aagacacgac ttatcgccac
1500

tggcagcagc cactggtaac aggattagca gagcgaggta tgttaggcggt gctacagagt
1560

10 tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttttgtt atctgcgc
1620

tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca
1680

ccgctggtag cggtggtttt tttgttgca agcagcagat tacgcgcaga aaaaaaggat
1740

20 ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggAAC gaaaactcac
1800

gttaaggat ttggcatg agattatcaa aaaggatctt cacctagatc cttttaaatt
1860

25 aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggct gacagttacc
1920

aatgcttaat cagtgaggca cctatctcgat cgatctgtct atttcgatca tccatagtt
30 1980

cctgactccc cgtcgtagataactacga tacgggaggg cttaccatct ggccccagtg
2040

35 ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca ataaaccagc
2100

cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc atccagtcta
2160

40 ttaatttggc ccgggaagct agagtaagta gttcgccagt taatagtttgcgcaacgttg
2220

ttgccattgc tacaggcatc gtgggtcac gctcgatcgat tggatggct tcattcagct
45 2280

ccggttccca acgatcaagg cgagttacat gatccccat gttgtcaaa aaagcggta
2340

50 gctccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcagtgatc tcactcatgg
2400

ttatggcagc actgcataat tctcttactg tcatgcccattc cgtaagatgc ttttctgtga
2460

55 ctggtgagta ctcaaccaag tcattctgag aatagtgtat gccccgaccg agttgctt
2520

gccccggcgtc aatacggat aataccgcgc cacatagcag aactttaaaa gtgctcatca
2580

5 ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt
2640

cgatgttaacc cactcgtgca cccaaactgat cttcagcata ttttactttc accagcgtt
2700

10 ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggataagg ggcacacgga
2760

aatgttgaat actcataactc ttccttttc aatattattt aagcatttat cagggttatt
15 2820

gtctcatgag cggatacata tttgaatgta ttttagaaaaaa taaacaaata ggggttccgc
2880

20 gcacattcc cgaaaaagtg ccacctg
2907

25 <210> 10

<211> 73

<212> DNA

30 <213> Artificial

35 <220>

<223> synthetic sequence

<220>

40 <221> sig_peptide

<222> (7) .. (69)

45 <400> 10
cgccatatga aaaagacagc tatcgcgatt gcagtggcac tggctggttt cgctaccgta
60

50 gcgcaagctt gag
73

55 <210> 11

<211> 21

<212> PRT

<213> Escherichia coli

5

<220>

10 <221> SIGNAL

<222> (1)..(21)

15

<400> 11

Met Lys Lys Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe Ala
1 5 10 15

20

Thr Val Ala Gln Ala
20

<210> 12

25

<211> 42

<212> DNA

30 <213> Artificial

<220>

35

<223> synthetic sequence

<400> 12

40 ggacatatgc tgaagcttcc caaccattc ccttatccag gc
42

<210> 13

45 <211> 28

<212> DNA

<213> Artificial

50

<220>

55 <223> synthetic sequence

<400> 13

cgcgatccc tccacagaga ggcactgc
28

5 <210> 14
<211> 578
<212> DNA
10 <213> Artificial

15 <220>
<223> synthetic sequence
<400> 14
20 catatgtga agctttccca accattccct tatccaggct ttttgacaac gctatgctcc
60
gccccatcg tctgcaccaag ctggccttg acacctacca ggagttgaa gaagcctata
120
25 tcccaaagga acagaagtat tcattcctgc agaacccttca gaccccttc tgtttctcag
180
30 agtctattcc gacaccctcc aacaggaggaa aacacaaca gaaatccaac ctagagctgc
240

tccgcatctc cctgctgctc atccagtcgt ggctggagcc cgtcagttc ctcaggagtg
300
35 tcttcgccaa cagcctggtg tacggcgct ctgacagcaa cgtctatgac ctcctaaagg
360

40 accttagagga aggcatccaa acgctgatgg ggaggctgga agatggcagc cccggactg
420

ggcagatctt caagcagacc tacagcaagt tcgacacaaa ctcacacaac gatgacgcac
480
45 tactcaagaa ctacgggctg ctctactgct tcaggaagga catggacaag gtcgagacat
540

50 tcctgcgcatt cgtcagtcg cgctctgtgg agggatcc
578

55 <210> 15
<211> 630
<212> DNA

<213> Artificial

5 <220>

<223> synthetic sequence

10 <220>

<221> source

<222> (4) .. (66)

15 <223> Eschericia coli

20 <220>

<221> source

<222> (67) .. (630)

25 <223> homo sapiens

30 <400> 15
catatgaaaa agacagctat cgcgattgca gtggcactgg ctggtttcgc taccgttagcg
60
caagcttcc caaccattcc cttatccagg ctttttgaca acgctatgct ccgcgcctat

35 120
cgtctgcacc agctggcctt tgacacctac caggagttt aagaagccta tatcccaaag
180
gaacagaagt attcattcct gcagaacccc cagacctccc tctgttctc agagtctatt
40 240
ccgacacctt ccaacaggga ggaaacacaa cagaaatcca acctagagct gctccgcattc
300
45 tccctgctgc tcatccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgccc
360
aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
420
50 gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tgggcagatc
480
55 ttcaaggcaga cctacagcaa gttcgacaca aactcacaca acgatgacgc actactcaag
540

aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
600

5 atcgtgcagt gccgctctgt ggagggatcc
630

<210> 16

10 <211> 208

<212> PRT

<213> Artificial
15

<220>

20 <223> synthetic sequence

<220>

<221> SIGNAL
25 <222> (1)..(21)

<223> from Eschericia coli

30

<400> 16

35 Met Lys Lys Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe Ala
1 5 10 15

Thr Val Ala Gln Ala Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp
20 25 30

40 Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr
35 40 45

Tyr Gln Glu Phe Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser
50 55 60

45 Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro
65 70 75 80

50 Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu
85 90 95

Leu Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln
100 105 110

55 Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp
115 120 125

Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr
130 135 140

Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe
5 145 150 155 160

Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala
165 170 175

10 Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp
180 185 190

Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly
195 200 205

15 <210> 17

<211> 570

20 <212> DNA

<213> Artificial

25 <220>

<223> synthetic sequence

30 <400> 17
catatgttcc caaccattcc cttatccagg cttttgaca acgctatgct ccgcggccat
60

35 cgtctgcacc agctggcctt tgacacctac caggagttt aagaaggcta tatcccaaag
120

gaacagaagt attcattcct gcagaacccc cagacctccc tctgtttctc agagtctatt
180

40 ccgacacccct ccaacaggga ggaaacacaa cagaaatcca accttagagct gctccgcata
240

tccctgctgc tcatccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgccc
300

45 aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
360

50 gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tgggcagatc
420

ttcaaggcaga cctacagcaa gttcgacaca aactcacaca acgatgacgc actactcaag
480

55 aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540

atcgtgcagt gccgctctgt ggagggatcc
570

5 <210> 18

<211> 188

<212> PRT
10 <213> Artificial

15 <220>

<223> synthetic sequence

<400> 18
20 Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
1 5 10 15

Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
25 20 25 30

Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
35 35 40 45

30 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
50 55 60

Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
65 70 75 80
35 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
85 90 95

Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
40 100 105 110

Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
115 120 125

45 Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
130 135 140

Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
145 150 155 160
50 Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
165 170 175

Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly
55 180 185

<210> 19
<211> 52
5 <212> DNA
<213> Artificial

10 <220>
<223> synthetic sequence

15 <400> 19
taccatatga catgatcatg tggcttcttc ccaaccattc ctttatccag gc
52

20 <210> 20
<211> 588

25 <212> DNA
<213> Artificial

30 <220>
<223> synthetic sequence

35 <400> 20
catatgacat gatcatgtgg cttttccca accattccct tatccaggct ttttgacaac
60

40 gctatgctcc gcccacatcg tctgcaccag ctggccttg acacccatca ggagtttcaa
120

gaaggctata tccaaagga acagaagtat tcattcctgc agaaccggca gacctccctc
180

45 tgtttcttag agtctattcc gacaccctcc aacaggggagg aaacacaaca gaaatccaaac
240

ctagagctgc tccgcacatc cctgctgctc atccagtcgt ggctggagcc cgtgcagttc
300

50 ctcaggaggta tcttcgcca cagcctggtg tacggcgcct ctgacagcaa cgtctatgac
360

55 ctcctaaagg accttagagga aggcatccaa acgctgatgg ggaggctgga agatggcagc
420

ccccggactg ggcagatctt caagcagacc tacagcaagt tcgacacaaa ctcacacaac
480

5 gatgacgcac tactcaagaa ctacggctg ctctactgct tcaggaagga catggacaag
540

588 gtcgagacat tcctgcgcac cgtgcagtgc cgctctgtgg agggatcc

10 10 <210> 21

<211> 191

15 15 <212> PRT

<213> Artificial

20 20 <220>

<223> synthetic sequence

25 25 <400> 21

Ser Cys Gly Phe Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn
1 5 10 15

30 30 Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr
20 25 30

Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe
35 40 45

35 35 Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr
50 55 60

40 40 Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu
65 70 75 80

Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe
85 90 95

45 45 Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser
100 105 110

Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu
115 120 125

50 50 Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys
130 135 140

Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu
55 145 150 155 160

Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys

165 170 175

Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly
180 185 190

5 <210> 22

 <211> 42

10 <212> DNA

 <213> Artificial

15 <220>

 <223> synthetic sequence

20 <400> 22
 tacgaattcc attgatcatg tggcttctag taggtcgacg at
 42

25 <210> 23

 <211> 4

 <212> PRT

30 <213> Artificial

35 <220>

 <223> synthetic sequence

40 <400> 23

40 Ser Cys Gly Phe
1

 <210> 24

45 <211> 51

 <212> DNA

50 <213> Artificial

55 <220>

 <223> synthetic sequence

<400> 24
tacgaattcc attgatcatg tggcttcaaa aagaaaatagt aggtcgacga t
51

5 <210> 25
<211> 7

10 <212> PRT
<213> Artificial

15 <220>
<223> synthetic sequence

20 <400> 25
Ser Cys Gly Phe Lys Lys Lys
1 5

25 <210> 26
<211> 1161
<212> DNA

30 <213> Artificial

35 <220>
<223> synthetic sequence

40 <400> 26
catatgacat gatcatgtgg cttcttccca accattccct tatccaggct ttttgacaac
60

gctatgctcc gcgcccatcg tctgcaccag ctggccttg acacctacca ggagttgaa
120

45 gaaggctata tcccaaagga acagaagtat tcattcctgc agaacccca gacccctc
180

50 ttttcctcag agtctattcc gacaccctcc aacaggggagg aaacacaaca gaaatccaac
240

ctagagctgc tccgcatttc cctgctgctc atccagtcgt ggctggagcc cgtgcagttc
300

55 ctcaggagtg tcttcgccaa cagcctggtg tacggcgct ctgacagcaa cgtctatgac
360

ctcctaaagg acctagagga aggcatcaa acgctgatgg ggaggctgga agatggcagc
420

5 ccccgactg ggcagatctt caagcagacc tacagcaagt tcgacacaaa ctcacacaac
480

gatgacgcac tactcaagaa ctacgggctg ctctactgct tcaggaagga catggacaag
540

10 gtcgagacat tcctgcgcat cgtcagtgcc cgctctgtgg agggatcatg tggcttc
600

ccaaccattc ctttatccag gcttttgac aacgctatgc tccgcgccca tcgtctgcac
15 660

cagctggcct ttgacaccta ccaggagttt gaagaagcct atatccaaa ggaacagaag
720

20 tattcattcc tgcaaaaccc ccagacctcc ctctgttct cagagtctat tccgacaccc
780

tccaacaggg agaaacaca acagaaatcc aacctagagc tgctccgcat ctccctgctg
840

25 ctcatccagt cgtggctgga gcccgtgcag ttccctcagga gtgtcttcgc caacagcctg
900

gtgtacggcg cctctgacag caacgtctat gacccctaa aggacctaga ggaaggcattc
30 960

caaacgctga tggggaggct ggaagatggc agccccggc ctggcagat cttcaaggcag
1020

35 acctacagca agttcgacac aaactcacac aacgatgacg cactactcaa gaactacggg
1080

ctgctctact gttcaggaa ggacatggac aaggtcgaga cattcctgcg catcgtgcag
1140

40 tgccgctctg tggaggatc c
1161

45 <210> 27

<211> 382

<212> PRT

50 <213> Artificial

55 <220>

<223> synthetic sequence

<400> 27

5 Ser Cys Gly Phe Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn
1 5 10 15

Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr
20 25 30

10 Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe
35 40 45

Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr
50 55 60

15 Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu
65 70 75 80

Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe
20 85 90 95

Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser
100 105 110

25 Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu
115 120 125

Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys
130 135 140

30 Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu
145 150 155 160

Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys
35 165 170 175

Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser
180 185 190

40 Cys Gly Phe Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala
195 200 205

Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln
210 215 220

45 Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu
225 230 235 240

Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro
50 245 250 255

Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg
260 265 270

55 Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu
275 280 285

Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn
290 295 300

5 Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met
305 310 315 320

Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln
325 330 335

10 Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu
340 345 350

Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val
355 360 365

15 Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly
370 375 380

20 <210> 28

<211> 1152

<212> DNA

25 <213> Artificial

30 <220>

<223> synthetic sequence

<220>

35 <221> misc_feature

<222> (574) .. (1146)

40 <223> sequence is repeated N-1 times, where N is a positive whole
numbe

45 <400> 28
tgatcatgtg gcttcttccc aaccattccc ttatccaggc ttttgacaa cgctatgctc
60

50 cgcgccatc gtctgcacca gctggcctt gacacctacc aggagttga agaaggctat
120

atcccaaagg aacagaagta ttcattcctg cagaaccccc agacccctt ctgtttctca
180

55 gagtctattc cgacaccctc caacagggag gaaacacaac agaaatccaa cctagagctg
240

ctccgcacatct ccctgctgct catccagtcg tggctggagc ccgtgcagtt cctcaggagt
300

5 gtcttcgcca acagcctggt gtacggcgcc tctgacagca acgtctatga cctcctaaag
360

gaccttagagg aaggcatcca aacgctgatg gggaggctgg aagatggcag cccccggact
420

10 gggcagatct tcaagcagac ctacagcaag ttgcacacaa actcacacaa cgatgacgca
480

ctactcaaga actacggct gctctactgc ttcaaggaagg acatggacaa ggtcgagaca
540

15 ttcctgcgca tcgtgcagtg ccgctctgtg gagggatcat gtggcttctt cccaaccatt
600

cccttatcca ggcttttga caacgctatg ctccgcgccc atcgctcgca ccagctggcc
20 660

tttgacacct accaggagtt tgaagaagcc tatatcccaa aggaacagaa gtattcattc
720

25 ctgcagaacc cccagacctc cctctgttcc tcagagtcta ttccgacacc ctccaacagg
780

gagggaaacac aacagaaatc caacctagag ctgctccgca tctccctgct gtcataccag
840

30 tcgtggctgg agcccggtgca gttcctcagg agtgtcttcg ccaacagcct ggtgtacggc
900

gcctctgaca gcaacgtctta tgacctccta aaggacctag aggaaggcat ccaaacgctg
35 960

atggggaggc tggaagatgg cagccccgg actgggcaga tcttcaagca gacctacagc
1020

40 aagttcgaca caaactcaca caacgatgac gcactactca agaactacgg gctgctctac
1080

tgcttcagga aggacatgga caaggtcgag acattcctgc gcatcggtca gtgccgctct
1140

45 gtggaggat cc
1152

50 <210> 29

<211> 382

<212> PRT

55 <213> Artificial

<220>
 5 <223> synthetic sequence
 <220>
 10 <221> MISC_FEATURE
 10 <222> (191) .. (381)
 15 <223> sequence is repeated N-1 times, where N is a positive whole
 numbe
 15
 <400> 29
 20 Ser Cys Gly Phe Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn
 1 5 10 15
 Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr
 25 20 25 30
 Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe
 35 40 45
 30 Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr
 50 55 60
 Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu
 65 70 75 80
 35 Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe
 85 90 95
 40 Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser
 100 105 110
 Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu
 115 120 125
 45 Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys
 130 135 140
 Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu
 145 150 155 160
 50 Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys
 165 170 175
 55 Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser
 180 185 190
 Cys Gly Phe Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala

	195	200	205	
	Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln			
	210	215	220	
5	Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu			
	225	230	235	240
	Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro			
10	245	250	255	
	Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg			
	260	265	270	
15	Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu			
	275	280	285	
	Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn			
	290	295	300	
20	Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met			
	305	310	315	320
	Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln			
25	325	330	335	
	Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu			
	340	345	350	
30	Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val			
	355	360	365	
	Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly			
	370	375	380	
35	<210> 30			
	<211> 606			
40	<212> DNA			
	<213> Artificial			
45	<220>			
	<223> synthetic sequence			
50	<400> 30			
	catatgttcc caaccattcc cttatccagg cttttgaca acgctatgct ccgcgccat			
	60			
	cgtctgcacc agctggcctt tgacacctac caggagtttg aagaagccta tatcccaaag			
55	120			

gaacagaagt attcattcct gcagaacccc cagacctccc tctgtttctc agagtctatt
180

ccgacacctt ccaacaggga ggaaacacaa cagaaatcca accttagagct gctccgcattc
5 240

tccctgctgc tcataccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgccc
300

10 aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
360

15 gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tgggcagatc
420

ttcaaggcaga cctacagcaa gttcgacaca aactcacaca acgtgacgc actactcaag
480

20 aactacgggc tgctctactg cttaggaag gacatggaca aggtcgagac attcctgcgc
540

atcgtgcagt gccgctctgt ggagggatcc gaattccatt gatcatgtgg cttctagtag
600

25 gtcgac
606

30 <210> 31

<211> 1737

<212> DNA

35 <213> Artificial

40 <220>

<223> synthetic sequence

<220>

45 <221> misc_feature

<222> (1138)..(1710)

50 <223> sequence is repeated N-1 times, where N is a positive whole
numbe

55 <400> 31

catatgttcc caaccattcc cttatccagg cttttgaca acgctatgct ccgcgcggat
60

5 cgtctgcacc agctggcctt tgacacctac caggagttt aagaagccta tatcccaaag
120

gaacagaagt attcattcct gcagaacccc cagacctccc tctgtttctc agagtctatt
180

10 ccgacaccct ccaacaggga ggaaacacaa cagaaatcca acctagagct gctccgcattc
240

tccctgtgc tcatccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgccc
300

15 aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
360

gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tgggcagatc
20 420

ttcaaggaga cctacagcaa gttcgacaca aactcacaca acgatgacgc actactcaag
480

25 aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540

atcgtgcagt gccgctctgt ggagggatca tgtggcttct tcccaaccat tcccttatcc
600

30 aggcttttg acaacgctat gctccgcgcc catcgctgc accagctggc ctttgacacc
660

taccaggagt ttgaagaagc ctatatccca aaggaacaga agtattcatt cctgcagaac
35 720

ccccagacct ccctctgttt ctcagagtct attccgacac cctccaaacag ggaggaaaca
780

40 caacagaaat ccaacctaga gctgctccgc atctccctgc tgctcatcca gtcgtggctg
840

gagccccgtgc agttcctcag gagtgttctc gccaacagcc tggtgtacgg cgcctctgac
900

45 agcaacgtct atgacctctt aaaggaccta gaggaaggca tccaaacgct gatggggagg
960

50 ctggaagatg gcagcccccg gactggcag atcttcaagc agacctacag caagttcgac
1020

acaaaactcac acaacgatga cgcaactactc aagaactacg ggctgctcta ctgcttcagg
1080

55 aaggacatgg acaaggcga gacattcctg cgcatcgatc agtgccgctc tgtggaggga
1140

tcatgtggct tcttcccaac cattccctta tccaggcttt ttgacaacgc tatgctccgc
1200

5 gcccatacgta tgccaccagct ggccttgac acctaccagg agtttgaaga agcctatatac
1260

ccaaaggAAC agaagtattc attcctgcag aaccccccaga cctccctctg tttctcagag
1320

10 tctattccga caccctccaa caggaggaa acacaacaga aatccaacct agagctgctc
1380

cgcatactccc tgctgctcat ccagtcgtgg ctggagccccg tgcagttcct caggagtgtc
1440

15 ttcgccaaca gcctggtgta cggcgctct gacagcaacg tctatgacct cctaaaggac
1500

20 ctagaggaag gcatccaaac gctgatgggg aggctggaag atggcagccc ccggactggg
1560

cagatctca agcagaccta cagcaagttc gacacaaact cacacaacga tgacgcacta
1620

25 ctcaagaact acgggctgct ctactgcttc aggaaggaca tggacaaggt cgagacattc
1680

ctgcgcacatcg tgcaagtgcgg ctctgtggag ggatcatgtg gttctagta ggtcgac
1737

30 <210> 32

35 <211> 574

<212> PRT

<213> Artificial

40 <220>

<223> synthetic sequence

45 <220>

<221> MISC FEATURE

50 <222> (379) .. (569)

<223> sequence is repeated N-1 times, where N is a positive whole
numbe

55

<220>
 <221> mat_peptide
 5 <222> (1)...()

 10 <400> 32
 Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
 1 5 10 15
 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
 15 20 25 30
 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 35 40 45
 20 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
 50 55 60
 Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
 65 70 75 80
 25 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
 85 90 95
 Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
 30 100 105 110
 Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
 115 120 125
 35 Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
 130 135 140
 Ser Lys Phe Asp Thr Asn Ser His Asn Asp Ala Leu Leu Lys Asn
 145 150 155 160
 40 Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
 165 170 175
 Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
 45 180 185 190
 Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg
 195 200 205
 50 Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu
 210 215 220
 Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro
 225 230 235 240
 55 Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg
 245 250 255

Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu
 260 265 270

5 Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val
 275 280 285

Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp
 290 295 300

10 Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu
 305 310 315 320

Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser
 15 325 330 335

Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr
 340 345 350

20 Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe
 355 360 365

Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe Phe
 25 370 375 380

Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala
 385 390 395 400

30 His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu
 405 410 415

Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln
 420 425 430

35 Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu
 435 440 445

Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu Leu
 40 450 455 460

Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe
 465 470 475 480

45 Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp Leu
 485 490 495

Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu
 50 500 505 510

Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser Lys
 515 520 525

Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly
 55 530 535 540

Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe Leu

545 550 555 560
Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
565 570

5 <210> 33
 <211> 55

10 <212> DNA
 <213> Artificial

15 <220>
 <223> synthetic sequence

20 <400> 33
taccatatga catgatcatg tggcttcggc ttcccaacca ttcccttatac caggc
55

25 <210> 34
 <211> 591
 <212> DNA
30 <213> Artificial

35 <220>
 <223> synthetic sequence
 <400> 34
40 catatgacat gatcatgtgg cttcggttcc ccaaccattc ccttatccag gcttttgac
60
 aacgctatgc tccgcgccca tcgtctgcac cagctggcct ttgacaccta ccaggagttt
120
45 gaagaagcct atatccaaa ggaacagaag tattcattcc tgcagaaccc ccagacctcc
180
 ctctgtttct cagagtctat tccgacacccc tccaacaggg aggaaacaca acagaaatcc
240
 aaccttagagc tgctccgcat ctccctgctg ctcatccagt cgtggctgga gccccgtgcag
300
55 ttcctcagga gtgtcttcgc caacagcctg gtgtacggcg cctctgacag caacgtctat
360

gacctcctaa aggacctaga ggaaggcatc caaacgctga tggggaggct ggaagatggc
420

5 agccccccga ctgggcagat cttcaagcag acctacagca agttcgacac aaactcacac
480

aacgatgacg cactactcaa gaactacggg ctgctctact gcttcaggaa ggacatggac
540

10 aaggtcgaga cattcctgctc catcgtagt tgccgctctg tggagggatc c
591

15 <210> 35

<211> 192

<212> PRT

20 <213> Artificial

25 <220>

<223> synthetic sequence

<400> 35

30 Ser Cys Gly Phe Gly Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp
1 5 10 15

Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr
20 25 30

35 Tyr Gln Glu Phe Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser
35 40 45

Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro
40 50 55 60

Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu
65 70 75 80

45 Leu Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln
85 90 95

Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp
100 105 110

50 Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr
115 120 125

Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe
55 130 135 140

Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala

145 150 155 160
Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp
165 170 175
5 Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly
180 185 190

<210> 36
10 <211> 1158

<212> DNA
15 <213> Artificial

20 <220>

<223> synthetic sequence

<220>
25 <221> misc_feature

<222> (577) ..(1152)

30 <223> sequence is repeated N-1 times, where N is a positive whole
 number

35 <400> 36
tgatcatgtg gcttcggttt cccaaaccatt cccttatcca ggcttttga caacgctatg
60

40 ctccgcgccc atcgcttgca ccagctggcc tttgacacct accaggagtt tgaagaagcc
120

 tatatcccaa aggaacagaa gtattcattc ctgcagaacc cccagacctc cctctgttcc
180
45 tcagagtcta ttccgacacc ctccaacagg gagaaacac aacagaaatc caaccttagag
240

50 ctgctccgca tctccctgct gctcatccag tcgtggctgg agcccgtgca gttcctcagg
300

 agtgttctcg ccaacagcct ggtgtacggc gcctctgaca gcaacgtcta tgacctccta
360
55 aaggacctag aggaaggcat ccaaacgctg atggggaggc tggaagatgg cagccccgg
420

actggcaga tcttcaagca gac tac a cgc a agt tcg a ca a act caca ca acg at gac
480

5 gcactactca agaactacgg gctgctcac tgcttcagga aggacatgga caaggcgag
540

acattcctgc gcatcgtgca gtgccgtct gtggaggat catgtggctt cggtttccca
600

10 accattccct tatccaggct tttgacaac gctatgctcc gcccacatcg tctgcaccag
660

ctggccttg acacccatcca ggagttgaa gaagcctata tcccaaagga acagaagtat
720

15 tcattcctgc agaaccacca gaccccttc tggatcttag agtctattcc gacaccctcc
780

20 aacaggagg aaacacaaca gaaatccaaac cttagagctgc tccgcacatc cctgctgctc
840

atccagtcgt ggctggagcc cgtgcagttc ctcaggagtg tcttcgcca cagcctggtg
900

25 tacggcgct ctgacagcaa cgtctatgac ctccctaaagg acctagagga aggcacccaa
960

acgctgatgg ggaggctgga agatggcagc cccggactg ggcagatctt caagcagacc
1020

30 tacagcaagt tcgacacaaa ctcacacaac gatgacgcac tactcaagaa ctacggctg
1080

35 ctctactgct tcaggaagga catggacaag gtcgagacat tccgcgcac cgtgcagtc
1140

cgctctgtgg agggatcc
1158

40 <210> 37

<211> 384

45 <212> PRT

<213> Artificial

50 <220>

<223> synthetic sequence

55 <220>

<221> MISC_FEATURE

<222> (192)...(383)

5 <223> sequence is repeated N-1 times, where N is a positive whole
numbe

10

<400> 37

Ser Cys Gly Phe Gly Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp
1 5 10 15

15

Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr
20 25 30

20

Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser
35 40 45

Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro
50 55 60

25

Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu
65 70 75 80

30

Leu Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln
85 90 95

Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp
100 105 110

35

Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr
115 120 125

Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe
130 135 140

40

Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala
145 150 155 160

Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp
165 170 175

45

Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly
180 185 190

50

Ser Cys Gly Phe Gly Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp
195 200 205

Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr
210 215 220

55

Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser
225 230 235 240

Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro
245 250 255

5 Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu
260 265 270

Leu Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln
275 280 285

10 Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp
290 295 300

Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr
305 310 315 320

15 Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe
325 330 335

Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala
20 340 345 350

Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp
355 360 365

25 Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly
370 375 380

<210> 38

30 <211> 1743

<212> DNA

35 <213> Artificial

<220>

40 <223> synthetic sequence

<220>

45 <221> misc_feature

<222> (1141)..(1716)

<223> sequence is repeated N-1 times, where N is a positive whole
50 numbe

<400> 38

55 catatgttcc caaccattcc cttatccagg ctttttgaca acgctatgct ccgcgcggcat
60

cgtctgcacc agctggcctt tgacacctac caggagttt aagaagccta tatcccaaag
120

5 gaacagaagt attcattcct gcagaacccc cagacctccc tctgttctc agagtctatt
180

ccgacaccct ccaacaggga ggaaacacaa cagaaatcca accttagagct gctccgcattc
240

10 tccctgtgc tcatccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgccc
300

aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
360

15 gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tggcagatc
420

20 ttcaagcaga cctacagcaa gttcgacaca aactcacaca acgtgacgc actactcaag
480

aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540

25 atcgtgcagt gccgctctgt ggagggatca tgtggcttcg gtttcccaac cattccctta
600

tccaggcttt ttgacaacgc tatgctccgc gcccattcg tcgaccagct ggccttgac
660

30 acctaccagg agtttgaaga agcctataatc ccaaaggaac agaagtattc attcctgcag
720

35 aaccccccaga cctccctctg tttctcagag tctattccga caccctccaa cagggaggaa
780

acacaacaga aatccaacct agagctgctc cgcatctccc tgctgctcat ccagtcgtgg
840

40 ctggagcccg tgcaaggccct caggagtgtc ttgcacaaca gcctgggtgt a cggcgccctct
900

gacagcaacg tctatgaccc cctaaaggac ctagaggaag gcatccaaac gctgatgggg
960

45 aggctgaaag atggcagccc ccggactggg cagatcttca agcagaccta cagcaagttc
1020

50 gacacaaact cacacaacga tgacgcacta ctcaagaact acgggctgct ctactgcttc
1080

aggaaggaca tggacaaggc cgagacattc ctgcgcattcg tgcagtgcgc ctctgtggag
1140

55 ggatcatgtg gcttcggttt cccaaaccatt cccttatcca ggcttttga caacgctatg
1200

ctccgcgcc 1260
atcgctgca ccagctggcc tttgacacctt accaggagtt tgaagaagcc
tatatccaa 1320
5 aggaacagaa gtattcattc ctgcagaacc cccagacctc cctctgttc
tcagagtcta 1380
10 ttccgacacc ctccaacagg gaggaaacac aacagaaaatc caaccttagag
ctgctccgca 1440
15 tctccctgct gctcatccag tcgtggctgg agcccgtgca gttcctcagg
agtgtttcg 1500
ccaacagcct ggtgtacggc gcctctgaca gcaacgtcta tgacccctta
aaggacctag 1560
1620
actggcaga tcttcaagca gacctacagc aagttcgaca caaactcaca caacgatgac
gcactactca 1680
agaactacgg gctgctctac tgcttcagga aggacatgga caaggtcgag
25 acattcctgc 1740
gcatcgtgca gtgccgctct gtggagggat catgtggctt ctagtaggtc
gac
1743
30
<210> 39
<211> 576
35 <212> PRT
<213> Artificial
40
<220>
<223> synthetic sequence
45 <220>
<221> MISC_FEATURE
50 <222> (380)..(571)
<223> sequence is repeated N-1 times, where N is a positive whole
numbe

55

<220>
 <221> mat_peptide
 5 <222> (1)...()

10 <400> 39
 Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
 1 5 10 15
 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
 15 20 25 30
 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 35 40 45
 20 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
 50 55 60
 Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
 25 65 70 75 80
 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
 85 90 95
 30 Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
 100 105 110
 Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
 115 120 125
 35 Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
 130 135 140
 Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
 40 145 150 155 160
 Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
 165 170 175
 45 Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
 180 185 190
 Gly Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
 195 200 205
 50 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
 210 215 220
 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 55 225 230 235 240
 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn

	245	250	255	
	Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser			
	260	265	270	
5	Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser			
	275	280	285	
10	Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr			
	290	295	300	
	Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg			
	305	310	315	320
15	Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr			
	325	330	335	
	Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn			
	340	345	350	
20	Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr			
	355	360	365	
	Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe			
25	370	375	380	
	Gly Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu			
	385	390	395	400
30	Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe			
	405	410	415	
	Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn			
35	420	425	430	
	Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn			
	435	440	445	
40	Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser			
	450	455	460	
	Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser			
	465	470	475	480
45	Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr			
	485	490	495	
	Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg			
50	500	505	510	
	Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr			
	515	520	525	
55	Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn			
	530	535	540	

Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
545 550 555 560

Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
5 565 570 575

<210> 40

<211> 39

10 <212> DNA

<213> Artificial

15 <220>

<223> synthetic sequence

20 <400> 40
cgcggatcct catgagaagc cacagctgcc ctccacaga
39

25 <210> 41

<211> 591

30 <212> DNA

<213> Artificial

35 <220>

<223> synthetic sequence

40 <400> 41
catatgttcc caaccattcc cttatccagg ctttttgaca acgctatgct ccgcgcctat
60

45 cgtctgcacc agctggcctt tgacacctac caggagttt aagaagccta tatcccaaag
120
gaacagaagt attcattcct gcagaacccc cagacctccc tctgtttctc agagtctatt
180

50 ccgacacccct ccaacaggga ggaaacacaa cagaaatcca accttagagct gctccgcata
240

55 tccctgctgc tcattccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgccc
300

aacagcctgg tgcacggcgc ctctgacagc aacgtctatg acctcctaaa ggaccttagag
360
gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tgggcagatc
5 420
ttcaagcaga cctacagcaa gttcgacaca aactcacaca acgatgacgc actactcaag
480
10 aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540
atcgtgcagt gccgctctgt ggagggcagc tgtggcttct catgaggatc c
591
15
<210> 42
<211> 193
20 <212> PRT
<213> Artificial
25
<220>
<223> synthetic sequence
30 <400> 42
Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
1 5 10 15
35 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
20 25 30
Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
40 35 40 45
Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
50 55 60
45 Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
65 70 75 80
Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
85 90 95
50 Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
100 105 110
Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
55 115 120 125
Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr

130 135 140
Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
145 150 155 160
5 Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
165 170 175
Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
10 180 185 190
Ser

15 <210> 43
 <211> 50
 <212> DNA
20 <213> Artificial

25 <220>
 <223> synthetic sequence
30 <400> 43
 catgccatgg ggtggtggag gaagttccc aaccattccc ttatccaggc
 50

35 <210> 44
 <211> 606
 <212> DNA
40 <213> Artificial

45 <220>
 <223> synthetic sequence
50 <400> 44
 ccatgggtg gtggaggaag tttcccaacc attcccttat ccaggcttt tgacaacgct
 60
 atgctccgcg cccatcgct gcaccagctg gccttgaca cctaccagga gtttgaagaa
 120
55 gcctatatcc caaaggaaca gaagtattca ttccctgcaga accccccagac ctccctctgt
 180

ttctcagagt ctattccgac accctccaac agggaggaaa cacaacagaa atccaaccta
240

5 gagctgctcc gcatctccct gctgctcatac cagtcgtggc tggagcccg gcagttcctc
300

aggagtgtct tcgccaacag cctggtgtac ggccgcctctg acagcaacgt ctatgacctc
360

10 ctaaaggacc tagaggaagg catccaaacg ctgatgggga ggctggaaga tggcagcccc
420

15 cggactgggc agatcttcaa gcagacctac agcaagttcg acacaaactc acacaacgat
480

gacgcactac tcaagaacta cgggctgctc tactgcttca ggaaggacat ggacaaggtc
540

20 gagacattcc tgcgcatcgt gcagtgccgc tctgtggagg gcagctgtgg cttctcatga
600

ggatcc
606

25 <210> 45

<211> 198

30 <212> PRT

<213> Artificial

35 <220>

<223> synthetic sequence

40 <400> 45

Trp Gly Gly Gly Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe
1 5 10 15

45 Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp
20 25 30

Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr
35 40 45

50 Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile
50 55 60

55 Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu
65 70 75 80

Leu Leu Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val

5	85	90	95
	Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser		
	100	105	110
10	Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln		
	115	120	125
15	Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile		
	130	135	140
	Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp		
	145	150	155
	Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met		
	165	170	175
20	Asp Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu		
	180	185	190
	Gly Ser Cys Gly Phe Ser		
	195		
25	<210> 46		
	<211> 603		
	<212> DNA		
30	<213> Artificial		
35	<220>		
	<223> synthetic sequence		
40	<400> 46 ccatggggtg gtggaggaag tttcccaacc attcccttat ccaggcttt tgacaacgcctt 60		
	atgctccgcg cccatcgct gcaccagctg gccttgaca cctaccagga gtttgaagaa 120		
45	gcctatatcc caaaggaaca gaagtattca ttccctgcaga acccccagac ctccccctgt 180		
	ttctcagagt ctattccgac accctccaac agggaggaaa cacaacagaa atccaaccta 240.		
50	gagctgctcc gcatctccct gctgctcatc cagtcgtggc tggagcccggt gcagttcctc 300		
	aggagtgtct tcgccaacag cctggtgtac ggcgcctctg acagcaacgt ctatgacctc 360		

ctaaaggacc tagaggaagg catccaaacg ctgatgggga ggctggaaga tggcagcccc
420

5 cggaactgggc agatcttcaa gcagacctac agcaagttcg acacaaaactc acacaacgt
480

10 gacgcactac tcaagaacta cgggctgctc tactgcttca ggaaggacat ggacaaggtc
540

15 gagacattcc tgcgcatcgt gcagtgcgc tctgtggagg gcagctgtgg cttcttaggga
600

tcc
603

15 <210> 47

20 <211> 197

<212> PRT

<213> Artificial

25 <220>

30 <223> synthetic sequence

<400> 47

Trp Gly Gly Gly Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe
1 5 10 15

35 Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp
20 25 30

40 Thr Tyr Gln Glu Phe Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr
35 40 45

Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile
50 55 60

45 Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu
65 70 75 80

Leu Leu Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val
85 90 95

50 Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser
100 105 110

55 Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln
115 120 125

Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile

130 135 140
Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp
145 150 155 160
5 Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met
 165 170 175
Asp Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu
10 180 185 190
Gly Ser Cys Gly Phe
 195
15 <210> 48
 <211> 1200
 <212> DNA
20 <213> Artificial

25 <220>
 <223> synthetic sequence
 <220>
30 <221> misc_feature
 <222> (595)..(1188)
35 <223> sequence is repeated N-1 times, where N is a positive whole
 numbe

40 <400> 48
 ccatgggtg gtggaggaag tttcccaacc attcccttat ccaggcttt tgacaacgct
 60
45 atgctcccg cccatcgct gcaccagctg gccttgaca cctaccagga gtttgaagaa
 120
 gcctatatcc caaaggaaca gaagtattca ttccctgcaga acccccagac ctccctctgt
 180
50 ttctcagagt ctattccgac accctccaac agggaggaaa cacaacagaa atccaaccta
 240
 gagctgctcc gcatctccct gctgctcatc cagtcgtggc tggagcccggt gcagttcctc
55 300

aggagtgtct tcgccaacag cctgggtac ggccgcctcg acagcaacgt ctatgacc
360

5 ctaaaggacc tagaggaagg catccaaacg ctgatgggg ggctggaaga tggcagccc
420

cggactggc agatctcaa gcagacctac agcaagttcg acacaaactc acacaacgat
480

10 gacgcactac tcaagaacta cgggctgctc tactgcttca ggaaggacat ggacaaggc
540

gagacattcc tgcgcatcgt gcagtgcgc tctgtggagg gcagctgtgg cttctcatgg
600

15 ggtggtgag gaagttccc aaccattccc ttatccaggg ttttgacaa cgctatgctc
660

cgcgccatc gtctgcacca gctggcctt gacacctacc aggagttga agaaggctat
720

20 atccccaaagg aacagaagta ttcattcctg cagaaccccc agacccctt ctgtttctca
780

25 gagtctattc cgacaccctc caacagggag gaaacacaac agaaatccaa cctagagctg
840

ctccgcacatct ccctgctgct catccagtcg tggctggagc cctgcagtt ctcaggagt
900

30 gtcttcgcca acagcctgggt gtacggcgcc tctgacagca acgtctatga ctcctaaag
960

35 gacctagagg aaggcatcca aacgctgatg gggaggctgg aagatggcag ccccccggact
1020

gggcagatct tcaagcagac ctacagcaag ttcgacaccaa actcacacaa cgatgacgca
1080

40 ctactcaaga actacggct gctctactgc ttcaggaagg acatggacaa ggtcgagaca
1140

ttcctgcgca tcgtgcagtg ccgcctgtg gagggcagct gtggcttctc atgaggatcc
1200

45 <210> 49

50 <211> 396

<212> PRT

<213> Artificial

55

<220>
 <223> synthetic sequence
 5 <220>
 <221> MISC_FEATURE
 10 <222> (198) .. (395)
 <223> sequence is repeated N-1 times, where N is a positive whole
 numbe
 15
 <400> 49
 Trp Gly Gly Gly Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe
 20 1 5 10 15
 Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp
 20 25 30
 25 Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr
 35 40 45
 Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile
 30 50 55 60
 Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu
 65 70 75 80
 Leu Leu Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val
 35 85 90 95
 Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser
 100 105 110
 40 Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln
 115 120 125
 Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile
 45 130 135 140
 Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp
 145 150 155 160
 50 Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met
 165 170 175
 Asp Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu
 55 180 185 190
 Gly Ser Cys Gly Phe Ser Trp Gly Gly Ser Phe Pro Thr Ile
 195 200 205

Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu
 210 215 220

5 His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile
 225 230 235 240

Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu
 10 245 250 255

Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln
 260 265 270

Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu Leu Ile Gln
 15 275 280 285

Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe Ala Asn Ser
 290 295 300

20 Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp
 305 310 315 320

Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu Asp Gly Ser
 325 330 335

25 Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr
 340 345 350

Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr
 30 355 360 365

Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe Leu Arg Ile Val
 370 375 380

35 Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe Ser
 385 390 395

<210> 50

40 <211> 1185

<212> DNA

<213> Artificial

45 <220>

50 <223> synthetic sequence

<400> 50
 catatgttcc caaccattcc cttatccagg ctttttgaca acgctatgct ccgcgcctat
 60

55 cgtctgcacc agctggcctt tgacacctac caggagttt aagaaggcta tatcccaaag
 120

gaacagaagt attcattcct gcagaacccc cagacacctcc tctgtttctc agagtctatt
180

5 ccgacaccct ccaacaggga ggaaacacaa cagaaatcca accttagagct gctccgcata
240

tccctgctgc tcataccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgcc
300

10 aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
360

15 gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tgggcagatc
420

ttcaagcaga cctacagcaa gttcgacaca aactcacaca acgtgacgc actactcaag
480

20 aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540

atcgtgcagt gccgctctgt ggagggcagc tgtggcttct catgggtgg tggaggaagt
600

25 ttcccaacca ttcccttatac caggctttt gacaacgcta tgctccgcgc ccatcgtctg
660

caccagctgg ccttgacac ctaccaggag ttgaagaag cctatatccc aaaggaacag
30 720

aagtattcat tcctgcagaa cccccagacc tcctctgtt tctcagagtc tattccgaca
780

35 ccctccaaca gggagggaaac acaacagaaa tccaacctag agctgctccg catctccctg
840

ctgctcatcc agtcgtggct ggagcccgta cagttcctca ggagtgtctt cgccaaacagc
900

40 ctggtgtacg ggcctctga cagcaacgta tatgacctcc taaaggacct agaggaaggc
960

atccaaacgc tggatggggag gctggaagat ggcagcccc ggactggca gatcttcaag
45 1020

cagacctaca gcaagttcga cacaaactca cacaacgatg acgcactact caagaactac
1080

50 gggctgctct actgcttcag gaaggacatg gacaaggctg agacattcct ggcgcattcgtg
1140

cagtgcgcct ctgtggaggg cagctgtggc ttctcatgag gatcc
1185

55

<210> 51
<211> 391
5 <212> PRT
<213> Artificial

10 <220>
<223> synthetic sequence

15 <220>
<221> mat_peptide
<222> (1)..()
20

<400> 51

25 Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
1 5 10 15
Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
20 25 30
30 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
35 40 45
Pro Gln Thr Ser Leu Cys Phe Ser Gln Ser Ile Pro Thr Pro Ser Asn
35 50 55 60
Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
65 70 75 80
40 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
85 90 95
Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
100 105 110
45 Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
115 120 125
Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
50 130 135 140
Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
145 150 155 160
55 Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
165 170 175

Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
180 185 190

5 Ser Trp Gly Gly Gly Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu
195 200 205

Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe
210 215 220

10 Asp Thr Tyr Gln Glu Phe Glu Ala Tyr Ile Pro Lys Glu Gln Lys
225 230 235 240

Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser
245 250 255

15 Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu
260 265 270

Glu Leu Leu Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro
20 275 280 285

Val Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala
290 295 300

25 Ser Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile
305 310 315 320

Gln Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln
325 330 335

30 Ile Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp
340 345 350

Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp
35 355 360 365

Met Asp Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val
370 375 380

40 Glu Gly Ser Cys Gly Phe Ser
385 390

<210> 52

45 <211> 1779

<212> DNA

<213> Artificial

50

<220>

55 <223> synthetic sequence

<220>
<221> misc_feature
5 <222> (1174) .. (1767)
<223> sequence is repeated N-1 times, where N is a positive whole
numbe
10

<400> 52
catatgttcc caaccattcc cttatccagg cttttgaca acgctatgct ccgcgcacat
15 60

cgtctgcacc agctggcctt tgacacctac caggagttt aagaagccta tatccaaag
120

20 gaacagaagt attcattcct gcagaacccc cagacctccc tctgtttctc agagtctatt
180

ccgacacccct ccaacaggga ggaaacacaa cagaaatcca acctagagct gctccgcata
240
25 tccctgctgc tcatccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgccc
300

aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagaa
360

gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tggcagatc
420

35 ttcaaggcaga cctacagcaa gttcgacaca aactcacaca acgtgacgc actactcaag
480

aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540
40 atcgtgcagt gcccgtctgt ggagggcagc tgtggcttct catgggtgg tggaggaagt
600

ttcccaacca ttcccttatac caggctttt gacaacgcta tgctccgcgc ccatcgtctg
45 660

caccagctgg ccttgacac ctaccaggag tttgaagaag cctatatccc aaaggaacag
720

50 aagtattcat tcctgcagaa cccccagacc tccctctgtt tctcagagtc tattccgaca
780

ccctccaaca gggagggaaac acaacagaaa tccaacctag agctgctccg catctccctg
840
55 ctgctcatcc agtctggct ggagccctg cagttcctca ggagtgctt cgccaacagc
900

ctgggtacg gcgcctctga cagcaacgtc tatgacctcc taaaggacct agaggaaggc
960

5 atccaaacgc tgatggggag gctggaagat ggcagcccc ggactggca gatcttcaag
1020

cagacctaca gcaagttcga cacaactca cacaacgtcg acgcactact caagaactac
1080

10 gggctgtct actgcttcag gaaggacatg gacaaggctcg agacattctt ggcacatcg
1140

cagtgccgct ctgtggaggg cagctgtggc ttctcatggg gtggtgagg aagtttccca
1200

15 accattccct tatccaggtt tttgacaac gctatgctcc ggcacccatcg tctgcaccag
1260

20 ctggcctttg acacctacca ggagttgaa gaagcctata tcccaaagga acagaagtt
1320

tcattcctgc agaaccccca gaccccttc tgtttcttag agtctattcc gacaccctcc
1380

25 aacagggagg aaacacaaca gaaatccaac ctagagctgc tccgcacatc cctgctgctc
1440

atccagtcgt ggctggagcc cgtcagttc ctcaggatgt tcttcgcca cagcctggtg
30 1500

tacggcgct ctgacagcaa cgtctatgac ctcctaaagg acctagagga aggcacatccaa
1560

35 acgctgatgg ggaggctgga agatggcagc cccggactg ggcagatctt caagcagacc
1620

tacagcaagt tcgacacaaa ctcacacaac gatgacgcac tactcaagaa ctacggctg
1680

40 ctctactgct tcaggaagga catggacaag gtcgagacat tcctgcgcac cgtgcagtgc
1740

45 cgctctgtgg agggcagctg tggctctca tgaggatcc
1779

<210> 53

50 <211> 589

<212> PRT

<213> Artificial

55

<220>

5 <223> synthetic sequence

10 <220>

15 <221> MISC_FEATURE

20 <222> (391) .. (588)

25 <223> sequence is repeated N-1 times, where N is a positive whole
numbe

30 15

35 <220>

40 <221> mat_peptide

45 <222> (1) .. ()

50 53

55 Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
1 5 10 15

60 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
20 25 30

65 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
35 40 45

70 35

75 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
50 55 60

80 Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
65 70 75 80

85 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
85 90 95

90 45

95 Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
100 105 110

105 Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
115 120 125

110 Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
130 135 140

115 55

120 Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
145 150 155 160

Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
 165 170 175
 5 Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
 180 185 190
 Ser Trp Gly Gly Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu
 195 200 205
 10 Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe
 210 215 220
 Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys
 225 230 235 240
 15 Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser
 245 250 255
 Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu
 260 265 270
 20 Glu Leu Leu Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro
 275 280 285
 25 Val Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala
 290 295 300
 Ser Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile
 305 310 315 320
 30 Gln Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln
 325 330 335
 Ile Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp
 340 345 350
 35 Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp
 355 360 365
 40 Met Asp Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val
 370 375 380
 Glu Gly Ser Cys Gly Phe Ser Trp Gly Gly Ser Phe Pro Thr
 385 390 395 400
 45 Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala His Arg
 405 410 415
 Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr
 420 425 430
 50 Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser
 435 440 445
 Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr
 450 455 460

Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu Leu Leu Ile
465 470 475 480

5 Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe Ala Asn
485 490 495

Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp Leu Leu Lys
500 505 510

10 Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu Asp Gly
515 520 525

Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser Lys Phe Asp
15 530 535 540

Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu
545 550 555 560

20 Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe Leu Arg Ile
565 570 575

Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe Ser
580 585

25 <210> 54

<211> 2370

30 <212> DNA

<213> Artificial

35 <220>

<223> synthetic sequence

40 <220>

<221> misc_feature

<222> (1174)..(1767)

45 <223> sequence is repeated N-1 times, where N is a positive whole
numbe

50 <400> 54
catatgttcc caaccattcc cttatccagg ctttttgaca acgctatgct ccgcggccat
60

55 cgtctgcacc agctggcctt tgacacctac caggagtttg aagaagccta tatcccaaag
120

gaacagaagt attcattcct gcagaacccc cagacccccc tctgtttctc agagtctatt
180

5 ccgacaccct ccaacaggga ggaaacacaa cagaaatcca acctagagct gctccgcata
240

tccctgctgc tcatccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgcc
300

10 aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
360

gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tggcagatc
15 420

ttcaagcaga cctacagcaa gttcgacaca aactcacaca acgtgacgc actactcaag
480

20 aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540

atcgtgcagt gccgctctgt ggagggcagc tgtggcttct catgggtgg tggaggaagt
600

25 ttcccaacca ttcccttatac caggctttt gacaacgcta tgctccgcgc ccatacgctg
660

caccagctgg cctttgacac ctaccaggag tttgaagaag cctatatccc aaaggaacag
30 720

aagtattcat tcctgcagaa cccccagacc tccctctgtt tctcagagtc tattccgaca
780

35 ccctccaaaca gggaggaaac acaacagaaa tccaaacctag agctgctccg catctccctg
840

ctgctcatcc agtcgtggct ggagcccttg cagttcctca ggagtgtctt cgccaaacagc
900

40 ctggtgtacg gcgcctctga cagcaacgta tatgacctcc taaaggacct agaggaaggc
960

atccaaacgc tgatggggag gctggaagat ggcagcccc ggactggca gatcttcaag
45 1020

cagacctaca gcaagttcga cacaaactca cacaacgatg acgcactact caagaactac
1080

50 gggctgctct actgcttcag gaaggacatg gacaaggctcg agacattcct ggcgcattcg
1140

cagtgcgcgt ctgtggaggg cagctgtggc ttctcatggg gtgggtggagg aagttccca
1200

55 accattccct tatccaggct ttttgcacac gctatgctcc ggcgcattcg tctgcaccag
1260

ctggccttg acacctacca ggagttgaa gaagcctata tcccaaagga acagaagtat
1320

5 tcattcctgc agaaccggca gaccccttc tgtttctcag agtctattcc gacaccctcc
1380

aacaggagg aaacacaaca gaaatccaac ctagagctgc tccgcatttc cctgctgctc
1440

10 atccagtcgt ggctggagcc cgtgcagttc ctcaggagtg tcttcgccaa cagcctggtg
1500

tacggcgct ctgacagcaa cgtctatgac ctcctaaagg acctagagga aggcatccaa
1560

15 acgctgatgg ggaggctgga agatggcagc ccccggactg ggcagatctt caagcagacc
1620

20 tacagcaagt tcgacacaaa ctcacacaac gatgacgcac tactcaagaa ctacggctg
1680

ctctactgct tcaggaagga catggacaag gtcgagacat tcctgcgcatt cgtgcagtgc
1740

25 cgctctgtgg agggcagctg tggcttctca tgggggtggtg gaggaagttt cccaaccatt
1800

cccttatcca ggcttttga caacgctatg ctccgcgccc atcgctgca ccagctggcc
1860

30 tttgacacct accaggagtt tgaagaagcc tatatccaa aggaacagaa gtattcattc
1920

35 ctgcagaacc cccagacctc cctctgttcc tcagagtcta ttccgacacc ctccaacagg
1980

gaggaacac aacagaaatc caaccttagag ctgctccgca tctccctgct gctcatccag
2040

40 tcgtggctgg agcccggtca gttcctcagg agtgtctcg ccaacagcct ggtgtacggc
2100

45 gcctctgaca gcaacgtcta tgacccctta aaggacctag aggaaggcat ccaaacgctg
2160

atggggaggc tggaaagatgg cagccccgg actgggcaga tcttcaagca gacctacagc
2220

50 aagttcgaca caaactcaca caacgatgac gcactactca agaactacgg gctgctctac
2280

tgtttcagga aggacatgga caaggtcgag acattcctgc gcatcgtca gtgccgctct
2340

55 gtggaggggca gctgtggctt ctagggatcc
2370

5 <210> 55
<211> 786
<212> PRT
10 <213> Artificial

15 <220>
<223> synthetic sequence
<220>

20 <221> MISC_FEATURE
<222> (391) .. (588)

25 <220>
<221> mat_peptide
30 <222> (1) .. ()

35 <400> 55
Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
1 5 10 15
Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
40 20 25 30
Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
35 40 45
45 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
50 55 60
Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
65 70 75 80
50 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
85 90 95
55 Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
100 105 110
Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg

	115	120	125
	Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr		
	130	135	140
5	Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn		
	145	150	155
	Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr		
10	165	170	175
	Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe		
	180	185	190
15	Ser Trp Gly Gly Gly Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu		
	195	200	205
	Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe		
	210	215	220
20	Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys		
	225	230	235
	Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser		
25	245	250	255
	Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu		
	260	265	270
30	Glu Leu Leu Arg Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro		
	275	280	285
	Val Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala		
	290	295	300
35	Ser Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile		
	305	310	315
	Gln Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln		
40	325	330	335
	Ile Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp		
	340	345	350
45	Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp		
	355	360	365
	Met Asp Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val		
	370	375	380
50	Glu Gly Ser Cys Gly Phe Ser Trp Gly Gly Gly Ser Phe Pro Thr		
	385	390	395
	Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala His Arg		
55	405	410	415
	Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr		

	420	425	430
	Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser		
	435	440	445
5	Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr		
	450	455	460
10	Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu Leu Leu Ile		
	465	470	475
	Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe Ala Asn		
	485	490	495
15	Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp Leu Leu Lys		
	500	505	510
	Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu Asp Gly		
	515	520	525
20	Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser Lys Phe Asp		
	530	535	540
25	Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu		
	545	550	555
	Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe Leu Arg Ile		
	565	570	575
30	Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe Ser Trp Gly Gly		
	580	585	590
	Gly Gly Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala		
	595	600	605
35	Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln		
	610	615	620
	Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu		
40	625	630	635
			640
	Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro		
	645	650	655
45	Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg		
	660	665	670
	Ile Ser Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu		
	675	680	685
50	Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn		
	690	695	700
	Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met		
55	705	710	715
			720
	Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln		

725 730 735

Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu
740 745 750

5 Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val
755 760 765

Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys
10 770 775 780

Gly Phe
785

15 <210> 56

<211> 33

<212> DNA
20 <213> Artificial

25 <220>

<223> synthetic sequence

30 <400> 56
33 ttaccatgga ttgccggcgg cggcgatcc aat
33

35 <210> 57

<211> 36

<212> DNA
40 <213> Artificial

45 <220>

<223> synthetic sequence

50 <400> 57
36 ttaccatgga tttgtatcagg cggcgccgga tccaa
36

55 <210> 58

<211> 36

<212> DNA
<213> Artificial
5
<220>
10 <223> synthetic sequence
<400> 58
tgatcaggcg gcggcggatc aggccgcggc ggatcc
36
15 <210> 59
<211> 10
20 <212> PRT
<213> Artificial

25 <220>
<223> synthetic sequence
30 <400> 59
Ser Gly Gly Gly Ser Gly Gly Gly
1 5 10
35 <210> 60
<211> 48
<212> DNA
40 <213> Artificial

45 <220>
<223> synthetic sequence
<400> 60
50 gcccggcg gcggtcagg cggccggcgta tcaggcggcg gcggtatcc
48

55 <210> 61
<211> 14

<212> PRT
<213> Artificial
5
<220>
10 <223> synthetic sequence
<400> 61
Gly Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly
1 5 10
15 <210> 62
<211> 43
20 <212> DNA
<213> Artificial
25 <220>
<223> synthetic sequence
30 <400> 62
ggacatatgc tgtgatcatt cccaaaccatt cccttatcca ggc
43
35 <210> 63
<211> 41
40 <212> DNA
<213> Artificial
45 <220>
<223> synthetic sequence
<400> 63
50 cgcgaattcg atccatggaa gccacagctg ccctccacag a
41
55 <210> 64
<211> 36

<212> DNA
<213> Artificial
5
<220>
10 <223> synthetic sequence
<400> 64
cgcgtcgacc tagaagccac agctgccctc cacaga
36
15 <210> 65
<211> 602
20 <212> DNA
<213> Artificial
25 <220>
<223> synthetic sequence
30 <400> 65
catatgtgt gatcattccc aaccattccc ttatccaggc ttttgacaa cgctatgctc
60
cgcgcccatc gtctgcacca gctggcctt gacacctacc aggagttga agaagcctat
35 120
atccccaaagg aacagaagta ttcattcctg cagaacccccc agacctccct ctgtttctca
180
40 gagtctattc cgacaccctc caacagggag gaaacacaac agaaatccaa cctagagctg
240
ctccgcatct ccctgctgct catccagtcg tggctggagc ccgtcagtt cctcaggagt
300
45 gtcttcgcca acagcctggt gtacggcgcc totgacagca acgtctatga ctcctaaag
360
50 gaccttagagg aaggcatcca aacgctgatg gggaggctgg aagatggcag ccccccggact
420
480 gggcagatct tcaagcagac ctacagcaag ttgcacacaa actcacacaa cgatgacgca
55 ctactcaaga actacggct gctctactgc ttgcaggaagg acatggacaa ggtcgagaca
540

ttcctgcgca tcgtgcagtgcgctctgtggggcagct gtggcttcca tggatcgaa
600

5 tc
 602

10 <210> 66

10 10 <211> 192

10 10 <212> PRT

15 15 <213> Artificial

15 15 <220>

20 20 <223> synthetic sequence

20 20 <400> 66

25 25 Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
 1 5 10 15

25 25 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
 20 25 30

30 30 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 35 40 45

35 35 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
 50 55 60

35 35 Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
 65 70 75 80

40 40 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
 85 90 95

40 40 Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
 100 105 110

45 45 Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
 115 120 125

50 50 Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
 130 135 140

50 50 Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
 145 150 155 160

55 55 Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
 165 170 175

55 55 Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe

180 185 190

5 <210> 67
5 <211> 600
5 <212> DNA
10 <213> Artificial
10

<220>
15 <223> synthetic sequence

<400> 67
catatgtgt gatcattccc aaccattccc ttatccaggg ttttgacaa cgctatgctc
60
20 cgccccatc gtctgcacca gctggcctt gacacctacc aggagttga agaaggctat
120

atcccaaagg aacagaagta ttcattcctg cagaacccccc agacccctt ctgtttctca
25 180

gagtctattc cgacaccctc caacagggag gaaacacaac agaaatccaa cctagagctg
240

30 ctccgcattct ccctgctgct catccagtcg tggctggagc ccgtcagtt cctcaggagt
300

gtttcgcca acagcctgggt gtacggcgcc tctgacagca acgtctatga ctcctaaag
360
35 gacctagagg aaggcatcca aacgctgatg gggaggctgg aagatggcag ccccccggact
420

40 ggcagatct tcaagcagac ctacagcaag ttcgacacaaa actcacacaa cgatgacgca
480

50 ctactcaaga actacggct gctctactgc ttcaggaagg acatggacaa ggtagagaca
540
45 ttcctgcgca tcgtgcagtg ccgctctgtg gagggcagct gtggcttcta ggtagacgct
600

<210> 68
50 <211> 192
50 <212> PRT
55 <213> Artificial

<220>

5 <223> synthetic sequence

<400> 68

10 Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
1 5 10 15

Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
20 25 30

15 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
35 40 45

Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
50 55 60

20 Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
65 70 75 80

25 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
85 90 95

Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
100 105 110

30 Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
115 120 125

Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
130 135 140

35 Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
145 150 155 160

40 Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
165 170 175

Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
180 185 190

45 <210> 69

<211> 639

50 <212> DNA

<213> Artificial

55 <220>

<223> synthetic sequence

<400> 69
catatgctgt gatcattccc aaccattccc ttatccaggc ttttgacaa cgctatgctc
60
5
cgccccatc gtctgcacca gctggcctt gacacctacc aggagttga agaaggctat
120
10
atcccaaagg aacagaagta ttcattcctg cagaaccccc agacctccct ctgttctca
180
15
gagtctattc cgacaccctc caacaggag gaaacacaac agaaatccaa cctagagctg
240
30
ctccgcatct ccctgctgct catccagtcg tggctggagc ccgtgcagtt cctcaggagt
360
20
gtcttcgcca acagcctgggt gtacggcgcc tctgacagca acgtctatga cctcctaaag
420
25
gacctagagg aaggcatcca aacgctgatg gggaggctgg aagatggcag ccccccggact
480
30
gggcagatct tcaagcagac ctacagcaag ttcgacacaa actcacacaa cgatgacgca
540
35
ctactcaaga actacggct gctctactgc ttcaggaagg acatggacaa ggtcgagaca
600
40
ttcctgcgca tcgtgcagtg ccgctctgtg gagggcagct gtggcttcgg cggcggcggaa
639
55
<210> 70
<211> 206
<212> PRT
<213> Artificial
45
<220>
50
<223> synthetic sequence
<400> 70
55
Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
1 5 10 15
Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
20 25 30

Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 35 40 45

5 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
 50 55 60

Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
 65 70 75 80

10 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
 85 90 95

Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
 15 100 105 110

Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
 115 120 125

20 Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
 130 135 140

Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
 145 150 155 160

25 Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
 165 170 175

Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
 30 180 185 190

Gly Gly Gly Ser Gly Gly Ser Gly Gly Gly Gly
 195 200 205

35 <210> 71

<211> 630

<212> DNA

40 <213> Artificial

45 <220>

<223> synthetic sequence

50 <400> 71
 catatgttcc caaccattcc cttatccagg ctttttgaca acgctatgct ccgcgcctat
 60

55 cgtctgcacc agctggcctt tgacacacctac caggagtttg aagaagccta tatcccaaag
 120

gaacagaagt attcattcct gcagaacccc cagacacctcc tctgtttctc agagtctatt
180

5 ccgacaccct ccaacaggga ggaaacacaa cagaaatcca acctagagct gctccgcata
240

tccctgtgc tcatccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgccc
300

10 aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
360

gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccccggac tgggcagatc
420

15 ttcaaggcaga cctacagcaa gttcgacaca aactcacaca acgtgacgc actactcaag
480

aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
20 540

atcgtgcagt gccgctctgt ggagggcagc tgtggcttcg gcggcggcgg atcaggcggc
600

25 ggcggatcag gcccggcgg atccgaattc
630

30 <210> 72

<211> 206

<212> PRT

35 <213> Artificial

40 <220>

<223> synthetic sequence

<400> 72

45 Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
1 5 10 15

Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
20 25 30

50 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
35 40 45

55 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
50 55 60

Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser

65 70 75 80
Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
85 90 95
5 Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
100 105 110
Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
10 115 120 125
Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
130 135 140
15 Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
145 150 155 160
Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
165 170 175
20 Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
180 185 190
25 Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Gly
25 195 200 205
<210> 73
<211> 1248
30 <212> DNA
<213> Artificial
35
<220>
<223> synthetic sequence
40 <220>
<221> misc_feature
45 <222> (619) .. (1236)
<223> sequence is repeated N-1 times, where N is a positive whole
 numbe
50
<400> 73
tgatcattcc caaccattcc cttatccagg ctttttgaca acgctatgct ccgcgcggcat
55 60

cgtctgcacc agctggcctt tgacacccat caggagtttg aagaagccta tatcccaaag
120

5 gaacagaagt attcattcct gcagaacccc cagacccccc tctgttctc agagtctatt
180

ccgacacccct ccaacaggga ggaaacacaa cagaaatcca accttagagct gctccgcata
240

10 tccctgctgc tcattccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgcc
300

aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
360

15 gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tggcagatc
420

20 ttcaaggcaga cctacagcaa gttcgacaca aactcacaca acgtgacgc actactcaag
480

aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540

25 atcgtgcagt gccgctctgt ggaggcgcgc tggggcttcg gcggcggcgg atcaggcgc
600

ggcggatcag gcggcggcgg atcattccca accattccct tatccaggct tttgacaac
660

30 gctatgctcc gcgcgcacatcg tctgcaccag ctggccttg acacccatca ggagttgaa
720

35 gaagcctata tcccaaagga acagaagtat tcattcctgc agaacccca gacccctc
780

tgtttctcag agtctattcc gacaccctcc aacaggagg aaacacaaca gaaatccaaac
840

40 ctagagctgc tccgcacatc cctgctgctc atccagtcgt ggctggagcc cgtgcagttc
900

ctcaggagtg tcttcgccaa cagcctggtg tacggcgcct ctgacagcaa cgtctatgac
960

45 ctcctaaagg acctagagga aggcatccaa acgctgatgg ggaggctgga agatggcagc
1020

50 ccccgactg ggcagatctt caagcagacc tacagcaagt tcgacacaaa ctcacacaac
1080

gatgacgcac tactcaagaa ctacggcgtc ctctactgct tcaggaagga catggacaag
1140

55 gtcgagacat tcctgcgcac cgtgcagtgc cgctctgtgg agggcagctg tggcttcggc
1200

ggcggccgat caggcggcgg cgatcaggc ggccggccgat ccgaattc
1248

5 <210> 74

<211> 412

<212> PRT
10 <213> Artificial

15 <220>

<223> synthetic sequence

<220>
20 <221> MISC_FEATURE

<222> (193) .. (398)

25 <223> sequence is repeated N-1 times, where N is a positive whole
numbe

30 <400> 74

Ser Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
1 5 10 15
35 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
20 25 30

Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
40 35 40 45

Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
50 55 60

45 Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
65 70 75 80

Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
85 90 95
50 Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
100 105 110

55 Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
115 120 125

Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr

	130	135	140
	Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn		
5	145	150	155
	Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr		
	165	170	175
10	Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe		
	180	185	190
	Gly Gly Gly Ser Gly Gly Ser Gly Gly Ser Phe		
	195	200	205
15	Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala		
	210	215	220
	His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu		
20	225	230	235
	240		
	Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln		
	245	250	255
25	Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu		
	260	265	270
	Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu Leu		
	275	280	285
30	Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe		
	290	295	300
	Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp Leu		
35	305	310	315
	320		
	Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu		
	325	330	335
40	Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser Lys		
	340	345	350
	Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly		
	355	360	365
45	Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe Leu		
	370	375	380
	Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe Gly Gly		
50	385	390	395
	400		
	Gly Gly Ser Gly Gly Ser Gly Gly Gly		
	405	410	
55	<210> 75		
	<211> 2445		

<212> DNA
<213> Artificial
5
<220>
10 <223> synthetic sequence
<220>
<221> misc_feature
15 <222> (1237) .. (1854)
<223> sequence is repeated N-1 times, where N is a positive whole
numbe
20
<400> 75
catatgttcc caaccattcc cttatccagg ctttttgaca acgctatgct ccgcgcacat
25 60
cgtctgcacc agctggcctt tgacacacctac caggagtttg aagaagccta tatcccaaag
120
30 gaacagaagt attcattcct gcagaacccc cagacctccc tctgtttctc agagtctatt
180
ccgacacccct ccaacaggga ggaaacacaaa cagaaatcca accttagagct gctccgcata
240
35 tccctgctgc tcatccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgcc
300
aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
40 360
gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tgggcagatc
420
45 ttcaaggcaga cctacagcaa gttcgacaca aactcacaca acgtgacgc actactcaag
480
aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540
50 atcgtgcagt gccgctctgt ggagggcagc tgtggcttcg gccccggcgg atcaggcggc
600
55 ggccgatcag gccccggcgg atcattccca accattccct tatccaggct ttttgacaaac
660

gctatgtcc gcgcccatcg tctgcaccag ctggccttg acacctacca ggagttgaa
720

5 gaaggctata tcccaaagga acagaagtat tcattcctgc agaacccca gacccctc
780

tgttcttag agtctattcc gacaccctcc aacaggagg aaacacaaca gaaatccaac
840

10 ctagagctgc tccgcacatc cctgctgctc atccagtcgt ggctggagcc cgtcagttc
900

ctcaggagtg tcttcgccaa cagcctggtg tacggcgct ctgacagcaa cgtctatgac
960

15 ctcctaaagg acctagagga aggcatccaa acgctgatgg ggaggctgga agatggcagc
1020

ccccggactg ggcagatctt caagcagacc tacagcaagt tcgacacaaa ctcacacaac
20 1080

gatgacgcac tactcaagaa ctacggctg ctctactgct tcaggaagga catggacaag
1140

25 gtcgagacat tcctgcgcac cgtcagtc cgcctgtgg agggcagctg tggcttcgac
1200

ggcggcggat caggcggcgg cgatcaggc ggcggcggat cattccaaac cattcccta
1260

30 tccaggcttt ttgacaacgc tatgctccgc gcccatcgac tgacccagct ggccttgac
1320

acctaccagg agtttaaga agcctataatc ccaaaggaac agaagtattc attcctgcag
35 1380

aaccccaaga cctccctctg tttctcagag tctattccga caccctccaa cagggaggaa
1440

40 acacaacaga aatccaacct agagctgctc cgcacatccc tgctgctcat ccagtcgtgg
1500

ctggagcccg tgcagttcct caggagtgtc ttgcacaaca gcctgggtta cggcgctct
1560

45 gacagcaacg tctatgacct cctaaaggac cttagaggaag gcatccaaac gctgatgggg
1620

50 aggctggaag atggcagccc ccggactggg cagatctca agcagaccta cagcaagttc
1680

gacacaaact cacacaacga tgacgcacta ctcaagaact acgggctgct ctactgcttc
1740

55 aggaaggaca tggacaaggt cgagacattc ctgcgcacatcg tgcagtgcgg ctctgtggag
1800

ggcagctgtg gcttcggcg cgccggatca ggccggcg gatcaggcg cggccgatca
1860

5 ttcccaacca ttcccttatac caggctttt gacaacgcta tgctccgcgc ccatcgctcg
1920

caccagctgg ccttgacac ctaccaggag tttgaagaag cctatatccc aaaggaacag
1980

10 aagtattcat tcctgcagaa cccccagacc tccctctgtt tctcagagtc tattccgaca
2040

ccctccaaca gggagggaaac acaacagaaa tccaacctag agctgctccg catctccctg
2100

15 ctgctcatcc agtcgtggct ggagccctgt cagttcctca ggagtgtctt cgccaaacagc
2160

ctggtgtacg gcgcctctga cagcaacgta tatgacacct taaaggacct agaggaaggc
20 2220

atccaaacgc tggatggggag gctggaagat ggcagcccc ggactggca gatcttcaag
2280

25 cagacctaca gcaagttcga cacaaactca cacaacgatg acgcactact caagaactac
2340

gggctgctct actgcttcag gaaggacatg gacaaggctcg agacattcct ggcgcattcg
2400

30 cagtgcgcct ctgtggaggg cagctgtggc ttcttaggtcg acgcg
2445

35 <210> 76

 <211> 810

 <212> PRT

40 <213> Artificial

45 <220>

 <223> synthetic sequence

 <220>

50 <221> MISC_FEATURE

 <222> (412) .. (617)

55 <223> sequence is repeated N-1 times, where N is a positive whole
 numbe

5 <220>
 <221> mat_peptide
 <222> (1)...()
 10
 <400> 76

15	Met Phe Pro Thr Ile Pro Leu Ser Arg	Leu Phe Asp Asn Ala Met Leu
	1 5	10 15
	Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr	Gln Glu Phe
	20 25	30
20	Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe	Leu Gln Asn
	35 40	45
	Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn	
	50 55	60
25	Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser	
	65 70	75 80
30	Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser	
	85 90	95
	Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr	
	100 105	110
35	Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg	
	115 120	125
	Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr	
	130 135	140
40	Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn	
	145 150	155 160
45	Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr	
	165 170	175
	Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe	
	180 185	190
50	Gly Gly Gly Ser Gly Gly Ser Gly Gly Gly Ser Phe	
	195 200	205
	Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala	
	210 215	220
55	His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu	
	225 230	235 240

	Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln		
	245	250	255
5	Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu		
	260	265	270
	Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu Leu		
	275	280	285
10	Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe		
	290	295	300
15	Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp Leu		
	305	310	315
	Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu		
	325	330	335
20	Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser Lys		
	340	345	350
	Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly		
	355	360	365
25	Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe Leu		
	370	375	380
30	Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe Gly Gly		
	385	390	395
	Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Ser Phe Pro Thr		
	405	410	415
35	Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala His Arg		
	420	425	430
	Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr		
	435	440	445
40	Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser		
	450	455	460
45	Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr		
	465	470	475
	Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu Leu Ile		
	485	490	495
50	Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe Ala Asn		
	500	505	510
	Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp Leu Leu Lys		
	515	520	525
55	Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu Asp Gly		
	530	535	540

Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser Lys Phe Asp
 545 550 555 560

5 Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu
 565 570 575

Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe Leu Arg Ile
 580 585 590

10 Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe Gly Gly Gly
 595 600 605

Ser Gly Gly Gly Ser Gly Gly Ser Phe Pro Thr Ile Pro
 15 610 615 620

Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His
 625 630 635 640

20 Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro
 645 650 655

Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys
 660 665 670

25 Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln
 675 680 685

Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu Leu Ile Gln Ser
 30 690 695 700

Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu
 705 710 715 720

35 Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu
 725 730 735

Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro
 740 745 750

40 Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn
 755 760 765

Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys
 45 770 775 780

Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe Leu Arg Ile Val Gln
 785 790 795 800

50 Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
 805 810

<210> 77

55 <211> 593

<212> DNA
<213> Artificial
5
<220>
<223> synthetic sequence
10 <400> 77
catatgttcc caaccattcc cttatccagg cttttgaca acgctatgct ccgcgcacat
60
15 cgtctgcacc agctggcctt tgacacctac caggagtttg aagaagccta tatcccaaag
120
gaacagaagt attcattcct gcagaacccc cagacctccc tctgttctc agagtctatt
180
20 ccgacacctt ccaacaggga gaaacacaa cagaaatcca acctagagct gctccgcac
240
25 tccctgctgc tcattccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgcc
300
aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggacctagag
360
30 gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tgggcagatc
420
ttcaaggcaga cctacagcaa gttcgacaca aactcacaca acgtacgc actactcaag
480
35 aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540
40 atcgtgcagt gccgctctgt ggagggcagc tgtggcttcc atggatcgaa ttc
593

<210> 78
45 <211> 192
<212> PRT
50 <213> Artificial

<220>
55 <223> synthetic sequence
<400> 78

Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
 1 5 10 15

5 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
 20 25 30

Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 35 40 45

10 Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
 50 55 60

15 Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
 65 70 75 80

Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser
 85 90 95

20 Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
 100 105 110

Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg
 115 120 125

25 Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr
 130 135 140

Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn
 30 145 150 155 160

Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr
 165 170 175

35 Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
 180 185 190

<210> 79

40 <211> 592

<212> DNA

<213> Artificial

45 <220>

50 <223> synthetic sequence

<400> 79
 aagctttccc aaccattccc ttatccaggc ttttgacaa cgctatgctc cgcgccccatc
 60

55 gtctgcacca gctggcctt gacacctacc aggagttga agaaggctat atcccaaagg
 120

aacagaagta ttcattcctg cagaacccccc agacccctt ctgtttctca gagtctatt
180

5 cgacaccctc caacaggag gaaacacaac agaaatccaa cctagagctg ctccgcac
240

ccctgctgct catccagtcg tggctggagc ccgtgcagtt cctcaggagt gtcttcgcca
300

10 acagcctggt gtacggcgcc tctgacagca acgtctatga ctcctaaag gacctagagg
360

15 aaggcatcca aacgctgatg gggaggctgg aagatggcag cccccggact gggcagatct
420

tcaaggcagac ctacagcaag ttgcacaccaa actcacaccaa cgatgacgca ctactcaaga
480

20 actacgggct gctctactgc ttccaggaagg acatggacaa ggtcgagaca ttccctgcga
540

25 tcgtgcagtg ccgctctgtg gagggcagct gtggcttcca tggatcgaat tc
592

<210> 80

<211> 191

30 <212> PRT

<213> Artificial

35 <220>

40 <223> synthetic sequence

<400> 80

Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg
1 5 10 15

45 Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu
20 25 30

Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro
50 35 40 45

Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg
50 55 60

Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu
65 70 75 80

Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val
85 90 95

5 Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp
100 105 110

Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu
115 120 125

10 Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser
130 135 140

Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr
145 150 155 160

15 Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe
165 170 175

Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
20 180 185 190

<210> 81

25 <211> 587

<212> DNA

<213> Artificial

30 <220>

35 <223> synthetic sequence

<400> 81
aagctttccc aaccattccc ttatccaggc ttttgacaa cgctatgctc cgcggccatc
60

40 gtctgcacca gctggcctt gacacctacc aggagttga agaaggctat atcccaaagg
120

aacagaagta ttcatccctg cagaaccccc agacccctt ctgtttctca gagtctattc
180

45 cgacaccctc caacagggag gaaacacaac agaaatccaa ccttagagctg ctccgcattc
240

ccctgctgct catccagtcg tggctggagc ccgtgcagtt cctcaggagt gtcttcgcca
300

acagcctgggt gtacggcgcc tctgacagca acgtctatga cctcctaaag gacctagagg
360

55 aaggcatcca aacgctgatg gggaggctgg aagatggcag ccccccggact gggcagatct
420

tcaaggcagac ctacagcaag ttgcacacaa actcacacaa cgtgacgca ctactcaaga
480

5 actacgggct gctctactgc ttccaggaagg acatggacaa ggtcgagaca ttccctgcgca
540

tcgtgcagtg ccgtctgtg gagggcagct gtggcttcta gggatcc
587

10 <210> 82

<211> 191

15 <212> PRT

<213> Artificial

20 <220>

<223> synthetic sequence

25 <400> 82

Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg
1 5 10 15

30 Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu
20 25 30

Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro
35 40 45

35 Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg
50 55 60

40 Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu
65 70 75 80

Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val
85 90 95

45 Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp
100 105 110

Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu
115 120 125

50 Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser
130 135 140

Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr
55 145 150 155 160

Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe

165 170 175

Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
180 185 190

5 <210> 83

 <211> 1165

10 <212> DNA

 <213> Artificial

15 <220>

 <223> synthetic sequence

20 <220>

 <221> misc_feature

 <222> (579) .. (1151)

25 <223> sequence is repeated N-1 times, where N is a positive whole
 numbe

30 <400> 83
 aagctttccc aaccattccc ttatccaggc ttttgacaa cgctatgctc cgcgccccatc
60

35 60
 gtctgcacca gctggcctt gacacctacc aggagttga agaaggctat atcccaaagg
120

40 aacagaagta ttcattcctg cagaacccccc agacaccttctca gagtttattc
180

 cgacaccctc caacagggag gaaacacaac agaaatccaa ccttagagctg ctccgcatct
240

45 ccctgctgct catccagtcg tggctggagc ccgtgcagtt cctcaggagt gtcttcgcca
300

50 acagcctgggt gtacggcgcc tctgacagca acgtctatga cctcctaaag gacctagagg
360

 aaggcatcca aacgctgatg gggaggctgg aagatggcag ccccccggact gggcagatct
420

55 tcaagcagac ctacagcaag ttcgacaccaa actcacacaa ccatgacgca ctactcaaga
480

actacgggct gctctactgc ttcaggaagg acatggacaa ggtcgagaca ttcctgcgca
540

5 tcgtgcagtg ccgctctgtg gagggcagct gtggcttctt cccaccatt cccttatcca
600

ggcttttga caacgctatg ctccgcgcc atcgctcgca ccagctggcc tttgacacct
660

10 accaggagtt tgaagaagcc tatatccaa aggaacagaa gtattcattc ctgcagaacc
720

cccagaccc tcctgtttc tcagagtcta ttccgacacc ctccaacagg gaggaaacac
780

15 aacagaaatc caaccttagag ctgctccgca tctccctgt gctcatccag tcgtggctgg
840

20 agcccggtgca gttcctcagg agtgtctcg ccaacagcct ggtgtacggc gcctctgaca
900

gcaacgtcta tgacctctta aaggacctag aggaaggcat ccaaacgctg atggggaggc
960

25 tggaaagatgg cagccccgg actgggcaga tcttcaagca gacctacagg aagttcgaca
1020

caaactcaca caacgatgac gcactactca agaactacgg gctgctctac tgttcagga
1080

30 aggacatgga caaggtcgag acattcctgc gcatcgatca gtgccgctct gtggagggca
1140

35 gctgtggctt ccatggatcg aattc
1165

40 <210> 84

<211> 191

<212> PRT

45 <213> Artificial

<220>

50 <223> synthetic sequence

<220>

<221> MISC_FEATURE

55 <222> (1) .. (191)

<223> sequence is repeated N times, where N is a positive whole number

5

<400> 84

Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu Arg

1 5 10 15

10

Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu
20 25 30

15

Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro
35 40 45

Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg
50 55 60

20

Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser Leu
65 70 75 80

Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val
85 90 95

25

Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr Asp
100 105 110

30

Leu Leu Lys Asp Leu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu
115 120 125

Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr Ser
130 135 140

35

Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr
145 150 155 160

Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr Phe
165 170 175

40

Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe
180 185 190

<210> 85

45

<211> 2307

<212> DNA

50

<213> Artificial

<220>

55

<223> synthetic sequence

<220>

5 <221> misc_feature

5 <222> (1153)..(1725)

5 <223> sequence is repeated N-1 times, where N is a positive whole
numbe

10

15 <400> 85
catatgttcc caaccattcc cttatccagg ctttttgaca acgctatgct ccgcgcggat
60

20 cgtctgcacc agctggcctt tgacacctac caggagtttg aagaaggcta tatcccaaag
120

25 gaacagaagt attcattcct gcagaacccc cagacctccc tctgtttctc agagtctatt
180

30 ccgacacccct ccaacaggga ggaaacacaa cagaaatcca accttagagct gctccgcata
240

35 tccctgctgc tcattccagtc gtggctggag cccgtgcagt tcctcaggag tgtcttcgccc
300

40 aacagcctgg tgtacggcgc ctctgacagc aacgtctatg acctcctaaa ggaccttagag
360

45 gaaggcatcc aaacgctgat ggggaggctg gaagatggca gccccggac tgggcagatc
420

50 ttcaaggcaga cctacagcaa gttcgacaca aactcacaca acgtgacgc actactcaag
480

55 aactacgggc tgctctactg cttcaggaag gacatggaca aggtcgagac attcctgcgc
540

60 atcgtgcagt gccgctctgt ggagggcagc tgggtttct tcccaaccat tcccttatcc
600

65 aggcttttg acaacgctat gctccgcgc catcgctctgc accagctggc ctttgacacc
660

70 taccaggagt ttgaagaagc ctatatccca aaggaacaga agtattcatt cctgcagaac
720

75 ccccagacct ccctctgttt ctcagagtct attccgacac cctccaacag ggaggaaaca
780

80 caacagaat ccaacctaga gctgctccgc atctccctgc tgctcatcca gtcgtggctg
840

85 gagcccggtgc agttcctcag gagtgtttc gccaacagcc tgggtacgg cgcctctgac
900

agcaacgtct atgacacctt aaaggaccta gaggaaggca tccaaacqct gatggggagg
960

5 ctgaaagatg gcagcccccg gactggcag atcttcaagc agacctacag caagttcgac
1020

10 acaaactcac acaacgatga cgcaactactc aagaactacg ggctgctcta ctgcttcagg
1080

15 aaggacatgg acaaggcga gacattcctg cgcatcgtgc agtgcgcgc tgtggagggc
1140

20 agctgtggct tcttcccaac cattccctta tccaggctt ttgacaacgc tatgctccgc
1200

25 gcccattcg tcgaccagct ggccttgac acctaccagg agtttgaaga agcctataatc
1260

30 ccaaaggaac agaagtattc attcctgcag aaccccccaga cctccctctg tttctcagag
1320

35 tctattccga caccctccaa cagggagggaa acacaacaga aatccaacct agagctgctc
1380

40 cgcattcccc tgctgctcat ccagtcgtgg ctggagcccg tgcagttcct caggagtgtc
1440

45 ttcgccaaca gcctggtgta cggcgccctt gacagcaacg tctatgacct cctaaaggac
1500

50 ctagaggaag gcatccaaac gctgatgggg aggctggaag atggcagccc ccggactggg
1560

55 cagatcttca agcagaccta cagcaagttc gacacaaact cacacaacga tgacgcacta
1620

60 ctcaagaact acgggctgct ctactgcttc aggaaggaca tggacaaggt cgagacattc
1680

65 ctgcgcattcg tgcagtgcgg ctctgtggag ggcagctgtg gcttctcccc aaccattccc
1740

70 ttatccaggc ttttgacaa cgctatgctc cggccccatc gtctgcacca gctggccctt
1800

75 gacacctacc agagatggta agaagcctat atcccaaagg aacagaagta ttcattcctg
1860

80 cagaacccccc agacctccct ctgtttctca gagtctattc cgacaccctc caacagggag
1920

85 gaaacacaac agaaatccaa cctagagctg ctccgcattt ccctgctgct catccagtcg
1980

90 tggctggagc ccgtgcagtt cctcaggagt gtcttcgcca acagcctgggt gtacggcgcc
2040

tctgacagca acgtctatga ctcctaaag gacctaggagg aaggcatcca aacgctgatg
2100

5 gggaggctgg aagatggcag ccccccggact gggcagatct tcaagcagac ctacagcaag
2160

ttcgacacaa actcacacaa cgatgacgca ctactcaaga actacggct gctctactgc
2220

10 ttcaggaagg acatggacaa ggtcgagaca ttctgcgcgca tcgtgcagtgc ccgctctgtg
2280

gagggcagct gtggcttcta gggatcc
15 2307

<210> 86

20 <211> 192

<212> PRT

<213> Artificial
25

<220>

30 <223> synthetic sequence

<220>

<221> MISC_FEATURE
35 <222> (2)..(192)

<223> sequence is repeated N+2 times, where N is a positive whole
numbe
40

<220>

45 <221> mat_peptide

<222> (1)..()

50

<400> 86

Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
55 1 5 10 15

Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe

	20	25	30
5	Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn 35	40	45
	Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn 50	55	60
10	Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser 65	70	75
	Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser 85	90	95
15	Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr 100	105	110
	Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg 115	120	125
20	Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln Thr Tyr 130	135	140
	Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn 145	150	155
25	Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp Lys Val Glu Thr 165	170	175
30	Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe 180	185	190

35

40

45